

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:55:10 ; Search time 42 Seconds
(without alignments)
2439.212 Million cell updates/sec

Title: US-09-804-357B-2

Perfect score: 2129
Sequence: 1 MSVGRRRVKLLGILMANVF.....RKQPMIDWISQLQSPNLKC 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP_ARCHAEA:*
- 2: SP_BACTERIA:*
- 3: SP_FUNGI:*
- 4: SP_HUMAN:*
- 5: SP_INVERTEBRATE:*
- 6: SP_MAMMAL:*
- 7: SP_MHC:*
- 8: SP_ORGANELLE:*
- 9: SP_PHAGE:*
- 10: SP_PLANT:*
- 11: SP_RODENT:*
- 12: SP_VIRUS:*
- 13: SP_VERTEBRATE:*
- 14: SP_UNCLASSIFIED:*
- 15: SP_VIRUS:*
- 16: SP_BACTERIAP:*
- 17: SP_ARCHAEA:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2119	99.5	397	11	Q91V18
2	2113	99.2	397	11	Q92222
3	1229.5	57.8	406	13	Q8UWM0
4	1176.5	55.3	412	13	Q8UWM3
5	688	32.3	412	13	Q8UWM2
6	658	30.9	397	11	Q8K437
7	655.5	30.8	401	4	Q8NF10
8	653.5	30.6	353	4	Q8N934
9	649	30.5	397	11	Q8KUJ2
10	648.5	30.5	388	4	Q8MD21
11	648	30.4	384	4	Q8TDX1
12	648	30.4	418	13	Q8UWM4
13	647.5	30.4	353	4	Q8WY02
14	647.5	30.4	353	4	Q8WWR5
15	647.5	30.4	378	4	Q9C0J1
16	647.5	30.4	378	4	Q8N5W4

17	643.5	30.2	353	4	Q96QH5
18	638.5	30.0	350	11	Q923H4
19	622.5	29.2	372	4	Q9C0J2
20	618	29.0	390	13	Q8UWM1
21	611.5	28.7	377	4	Q96EK0
22	608	28.6	372	11	Q8R0U2
23	605	28.4	372	11	Q9D722
24	602	28.3	370	11	Q8BK98
25	602	28.3	399	11	Q8V116
26	569.5	26.7	277	4	Q8TA24
27	565	26.5	389	11	Q8R319
28	524.5	24.6	374	4	Q8WR86
29	480.5	22.6	326	4	Q91526
30	480.5	22.6	326	6	Q9MYM7
31	480.5	22.6	326	11	Q54904
32	470	22.1	305	11	Q91V52
33	470	22.1	305	11	Q920V5
34	427.5	20.1	378	4	Q8BYG0
35	408	19.2	409	11	Q920V2
36	405	19.0	409	11	Q91V58
37	405	19.0	422	11	Q8CBX4
38	405	19.0	422	11	Q8BH19
39	404	19.0	409	11	Q91VE9
40	403.5	19.0	376	11	Q8BGY6
41	402	18.9	409	11	Q920V4
42	401	18.8	409	11	Q91V19
43	401	18.8	409	11	Q920V3
44	400.5	18.8	422	11	Q54905
45	400	18.8	422	4	Q43825

ALIGNMENTS

RESULT 1

Q91V18					
ID	Q91V18	PRELIMINARY;	PRT;	397	AA.
AC	Q91V18				
DT	01-DEC-2001 (TRENBLrel. 19, Created)				
DT	01-DEC-2001 (TRENBLrel. 19, Last sequence update)				
DT	01-OCT-2002 (TRENBLrel. 22, Last annotation update)				
DE	Beta-1,3-N-acetylglucosaminyltransferase (Beta-1,3-N-acetylglucosaminyltransferase 1).				
DE	B3GNT1.				
GN	Mus musculus (Mouse).				
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C.B17;				
RA	Egan S.E., Cohen B.L., Sarkar M., Ying Y., Cohen S., Singh N.,				
RA	Wang W., Flock G., Goh T., Schachter H.;				
RT	*Molecular cloning and expression analysis of a mouse UDP-				
RT	GlcNAc:Gal(beta1-4)Glc(NAC)-R beta1,3-N-acetylglucosaminyltransferase				
RT	homologous to Drosophila melanogaster Brainiac and the beta1,3-				
RT	galactosyltransferase family*;				
RL	Glycocon]. J. 17:865-872(2000).				
RL	[2]				
RN	SEQUENCE FROM N.A.				
RP	STRAIN=C.B17;				
RC	STRAIN=C.B17;				
RA	Egan S.E., Cohen B.L., Sarkar M., Ying Y., Cohen S., Singh N.,				
RA	Wang W., Flock G., Goh T., Schachter H.;				
RT	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.				
RL	[3]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Breast tumor;				
RA	Strausberg R.;				
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AY043479; AAK95359.1; -				
DR	EMBL: BC009075; AAH09075.1; -				
DR	MGI: 1889505; B3gnt1.				
DR	InterPro: IPR002659; Glyco_trans_31.				

Q96QH5	homo sapien
Q923H4	mus musculus
Q9C0J2	homo sapien
Q8UWM1	brachydanio
Q96EK0	homo sapien
Q8R0U2	mus musculus
Q9D722	mus musculus
Q8BK98	mus musculus
Q8V116	mus musculus
Q8TA24	homo sapien
Q8R319	mus musculus
Q8WR86	homo sapien
Q91526	homo sapien
Q9MYM7	pongo pygma
Q54904	mus musculus
Q91V52	mus musculus
Q920V5	mus spicile
Q9BYG0	homo sapien
Q920V2	mus spicile
Q91V58	mus musculus
Q8CBX4	mus musculus
Q8BH19	mus musculus
Q91VE9	mus musculus
Q8BGY6	mus musculus
Q920V4	mus musculus
Q91V19	mus musculus
Q920V3	mus musculus
Q54905	mus musculus
Q43825	homo sapien

```
DR Pfam: PF01762; Galactosyl_T; 1.
KW Glycosyltransferase: Transferase.
SQ SEQUENCE 397 AA: 45883 MW: D8B8EAL866CID106 CRC64;

Query Match          99.5%; Score 2119; DB 11; Length 397;
Best Local Similarity 99.5%; Pred. No. 3.1e-186;
Matches 395; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSVGRRRVLLGILMMANVFYILIVEVSKNSQDKNGKGVIIIPKEKFKWKPSTPRAYWN 60
DB 1 MSVGRRRVLLGILMMANVFYILIVEVSKNSQDKNGKGVIIIPKEKFKWKPSTPRAYWN 60
QY 61 REQELNRWYNPILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDFL 120
DB 61 REQELNRWYNPILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDFL 120
QY 121 LYLRCRNYSLIDQPKKCAKPFLLAIKSLIPHARQAIRRESGRTNNGTIVRVF 180
DB 121 LYLRCRNYSLIDQPKKCAKPFLLAIKSLIPHARQAIRRESGRTNNGTIVRVF 180
QY 181 LLGKTPPEDNHPDLSMDLKFESDKHQDILMWNRYRDTFFNLKVEYLFWRWYSTSCPDAEF 240
DB 181 LLGKTPPEDNHPDLSMDLKFESDKHQDILMWNRYRDTFFNLKVEYLFWRWYSTSCPDAEF 240
QY 241 VFKGDDVFNTHHILNLSLSKSKAKDLFGDVVHGNAGPHRDKKLYIPEVFTGYV 300
DB 241 VFKGDDVFNTHHILNLSLSKSKAKDLFGDVVHGNAGPHRDKKLYIPEVFTGYV 300
QY 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLGVPEKHKGFRTEIDEE 360
DB 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLGVPEKHKGFRTEIDEE 360
QY 361 KNKKNICSVIDMLVHSRKPQEMIDWSQSPNLKC 397
DB 361 KNKKNICSVIDMLVHSRKPQEMIDWSQSPNLKC 397

RESULT 2
Q92222
ID Q92222 PRELIMINARY; PRT: 397 AA.
AC Q92222;
DT 01-MAR-1999 (Tremblrel. 10, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE N-acetylglucosaminide beta-1,3-N-acetylglucosaminyltransferase
DE (EC 2.4.1.149) (Poly-N-acetylglucosamine extension enzyme) (N-
DE acetylglucosaminyltransferase) (Beta3GNT).
GN B3GNT1 OR BETA3GNT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN-ICR; TISSUE-Neonatal brain;
RX MEDLINE=99110903; PubMed=9892646;
RA Zhou D., Dinter A., Gutierrez Gallego R., Kamerling J.P.,
RA Vliegdenhart J.F.G., Berger E.G., Hennet T.;
RT "A beta-1,3-N-acetylglucosaminyltransferase with poly-N-
RT acetylglucosamine synthase activity is structurally related to beta-
RT 1,3-galactosyltransferases.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:406-411(1999).
RN [2]
RP REVISIONS.
RA Zhou D., Berger E.G., Hennet T.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CAN INITIATE THE SYNTHESIS OR THE ELONGATION OF THE
CC LINEAR POLY-N-ACETYLGLUCOSAMINOGLYCANS.
CC -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-
CC GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-
CC GLUCOSAMINYL-1,3-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.
CC -!- COFACTOR: MANGANESE.
CC -!- PATHWAY: GLYCOSYLATION.
```



```

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF321831; AAL32299.1; -.
DR InterPro: IPR002659; Glyco_trans_31.
DR Pfam: PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 406 AA; 46953 MW; F928471BBDF1681 CRC64;

Query Match 57.8%; Score 1229.5; DB 13; Length 406;
Best Local Similarity 57.9%; Pred. No. 1.8e-104;
Matches 226; Conservative 63; Mismatches 98; Indels 3; Gaps 2;

Qy 5 RRRVLLGLILMANVFIYLIVESKNSOOKNGKGGVILPKSEKFWKPPSTPRAYNRRQE 64
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 5 RRVKVKMA--MMTWVFLFVVEVSRRAGKSSKNSLVLKRFKAKDLPSDAYNRRQQ 62
Qy 65 KLRWYNPILNRVANOTGELATSPN-TSHLSYCEPDSTVMTAVTDFNNLPDRFKDPLLYL 123
Dy :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Dy 63 QINYNRNLEKYNIDNLPWLNDTVSLDSCDPDYRVTTVKDINSLPDRFKDPLLYM 122
Qy 124 RCNRYSLLDIDOPKKCAKPFLLAIAKSLIPHFAARRQAIRESGRETNGVTVVRVFLG 183
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 123 RCRSYPIVVDPNICKQKPFLLAIAKSLVPHFDRRQAIRESGKVGRIANRSVWTVLLG 182
Qy 184 KTPPEONHDPDLSOMLKFESEKHODILMNYRDTFFNLKKEVLFLWVSTPCDPAEFVK 243
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 183 NAATEDHFPDLSKMLHSESHIRDILOWDYRDTFFNLTKKEVLFLWVSTPCPGANFIK 242
Qy 244 GDDOVFNTHHILNLSLSKSKAKOLFIDGVTHNAGPHRDKKLYYIPEVFTGVYPPY 303
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 243 GDDOVFNTHIIDLNLNSNAKARELFVGDVITNAGPHRDKKLYYIPESMFVGMYPAY 302
Qy 304 AGGGGFLYSPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHGKFRFTDIEENK 363
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 303 AGGGGFLFSQLAQRUNTSKLYPLYPIDDVYTGMCLOKLGVLPEKHGKFRFTDIEEYR 362
Qy 364 KNICSYIDLMLVHSRKPQEMIDIWSQLSQSP 393
Dy ||: ||: |||||: || |||||: |||||: |||||: |||||: |||||: |||||:
Dy 363 DNACAYKSLMLVHPRSPQHMKIWMLNDP 392

RESULT 4
Q8UWM3 PRELIMINARY; PRT; 412 AA.
AC Q8UWM3 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Beta-3-galactosyltransferase.
GN SSP2.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Chou C.-M., Leu J.-H., Huang C.-J.;
RT *A family of novel genes encoding beta-3-galactosyltransferase from
RT zebrafish, zssp3 mRNA.*;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF321828; AAL32296.1; -.
DR InterPro: IPR002659; Glyco_trans_31.
DR Pfam: PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 412 AA; 47440 MW; 14F24C26570B5F45 CRC64;

Query Match 55.3%; Score 1176.5; DB 13; Length 412;
Best Local Similarity 57.1%; Pred. No. 1.4e-99;
Matches 230; Conservative 55; Mismatches 105; Indels 13; Gaps 6;

Qy 1 MSVGRRRVKLLGILMANVFIYLIVESKNSODKNG-KGGVILPKSEKFWKPPSTPRAYW 59
Dy |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 1 MSNNWNTKLLGLMLANFLIYIAVESRSHNDRNAFKKNKHIPKE-FWKKIKYSTAFW 59

60 NREOEKLNLR-WYNPILNRVANOTGELATS-----PNTSHLSYCEPDSTVMTAVTDFNNLPD 114
Dy |||||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
60 NREOEKLDIYIPLVN-----ASELPPHPRGIPRLNHS-CKHDVNVTAQIEDFNLSLPD 113
Qy 115 RFKDFLLYLRNRYSLLDIDOPKKCAKPFLLAIAKSLIPHFAARRQAIRESGRETNGVNGQ 174
Dy |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
114 RFQDFLLYMGCRSYPLITRAPKVCSPPPYLLAIAKSLAPHFDRRQAIRESGWRAGILDGO 173
Qy 175 TVVRVFLGKTPEDNHPDLSOMLKFESEKHODILMNYRDTFFNLKKEVLFLRWVST 234
Dy |||||: ||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
174 RIATVFVLGNTASTDHPDLSNMVYKHEALYGLQWLDYRDTFFNLTKKEVLFLWFGSH 233
Qy 235 CPDAEFVFKGDDOVFNTHHILNLSLSKSKAKOLFIDGVTHNAGPHRDKKLYYIPEV 294
Dy |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
234 CASAQYVFKGDDOVFNTRHMLAYLANFSISSQDLEFIGDVITNAGPHRSRLKYIYPE 293
Qy 295 FYTGVPYPPYAGGGFLYSPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHGKFR 354
Dy |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
294 VFTGGYPPYAGGGFLYSGVGLRLKISRLVLYPIDDVYTGMCLOKLGVLPEKHGKFR 353
Qy 355 TFDTEENKKNICSYIDLMLVHSRKPQEMIDIWSQLSQSPNLC 397
Dy |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
354 TFDIEAKHRENPAYKGLILDVHPRSPQDMIKIWSWINDPNAIC 396

RESULT 5
Q8UWM2 PRELIMINARY; PRT; 412 AA.
AC Q8UWM2 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Beta-3-galactosyltransferase.
GN SSP3.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Chou C.-M., Leu J.-H., Huang C.-J.;
RT *A family of novel genes encoding beta-3-galactosyltransferase from
RT zebrafish, zssp3 mRNA.*;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF321829; AAL32297.1; -.
DR InterPro: IPR002659; Glyco_trans_31.
DR Pfam: PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 412 AA; 47812 MW; 80C554675A2CB00A CRC64;

Query Match 32.3%; Score 688; DB 13; Length 412;
Best Local Similarity 37.2%; Pred. No. 1.1e-54;
Matches 135; Conservative 76; Mismatches 124; Indels 28; Gaps 5;

Qy 60 NREOEKLNRYNPIILNRVANOTGELATSPNTSHLS-----YCEPD 99
Dy |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
47 DKYVEVGRMCGP--QCPSPFRKNLKAVENTSSGSDSKRAFKPLPKKWDVNTCTEN 104
Qy 100 STVMYATVDFNNLPDRFKDILYLRNRYSLLDIDOPKKCAKPFLLAIAKSLIPHFAARRQ 159
Dy ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
105 STIKTQLW-FRRLSPRPFVLRHRCRYFPMLLNHPKCGGGVDVLVVYVKSVEIHDRRE 163
Qy 160 AIRESGRETNGVNTVVRVFLGKTPEDNHPDLSOMLKFESEKHODILMNYRDTFFN 219
Dy |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
164 AVRTWKGQEIQGLKIKTKTLFLGTFAPGKDSRNLAQALVQYEDRTYGDILQWDFMDTFEN 223
Qy 220 LSLKEVLFLRWVSTSCDPAEFVKGDDVFNTHHILNLSLSKSKAKOLFIDGVTHN 278
Dy |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
224 LTLKEVFLRWVFSYCPDVPFIKGGDDVFPVHTKNLVELIGFRKEENKVENLIVDAILE 283
Qy 279 AGPHRDKKLYYIPEVFTGVYPPYAGGGFLYSGFALLRLYSATSRVHLYPIDDVYTGW 338
```



```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Leu J.H., Chou C.M., Huang C.J.:
RT "Cloning and expression of a novel human beta-1.3-
RT galactosyltransferase-related gene.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF321825; AAL37219.1;
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 353 AA; 39847 MW; FB91097D5120AC50 CRC64;

Query Match 30.4%; Score 647.5; DB 4; Length 353;
Best Local Similarity 39.7%; Pred. No. 4.7e-51;
Matches 147; Conservative 53; Mismatches 115; Indels 55; Gaps 11;

QY 40 GVIIPEKFKWPPSTPRA---YNNREQEKLNRWYNPILNRVANQTGELATSPNTSHLSYC 96
DB 19 GCLLFVRKTAAGDPTAHQPFW-----APTPRHRSRC 51
QY 97 EPDSTVMTAVTDNNLPDRFKDFLLYLCRCNYSLLIDQPKKCAKKPFLLLAIKSLIPHEA 156
DB 52 PPNHTVSSASL---SLPSRHLFLTYRHCNFSILL-EPSCGSKDTFLLAIAKSPGHVE 107
QY 157 RQAIRESWGRETNVG-----NQTVVVRVFL---GKTPPEDNHPDLSMDMLKFESDKHQDIL 209
DB 108 RRAAIRSTWGR---VGGWARGROLKLVFLGVAGSAPP-----AQLLAYESREFDDIL 157
QY 210 MNRYRTFFNLISLKEVLFRLWVSTSCPDAEFVKGDGDDVFVNTHILNLSLSKSKAKD 269
DB 158 QWDETEDEFFNLTKELHLQRWVVAACPAHFMLKGGDDVFVHPNVLEFDGWN--DPAQD 215
QY 270 LFGDVTNAGPRHKKLYIPEVFTGV-YPPYAGGGGFLYSGPALLRLYSATSRRVHL 328
DB 216 LLVGDTVIRQALPNRNTKVYFIPPSMYRATHYPPYAGGGGVYMSRATVRRLOAIMEDAEL 275
QY 329 YPIDDVYTGMCLOKLGVLPEKHKGFRTFDEEK-NKNKICSYIDLMLVHSRKPQEMIDIW 387
DB 276 FPIDDVFGVGMCLRLGLSPMHAGFKTFGIRRLDPLDPCLYRGLLLVHRLSPLEMTWM 335
QY 388 SLOQSPNLKC 397
DB 336 ALVTDEGLKC 345

RESULT 14
Q8WWR5 PRELIMINARY; PRT; 353 AA.
AC Q8WWR5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Beta 1.3 galactosyltransferase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bennett E.P.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ278962; CAC82375.1;
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 353 AA; 39863 MW; EAA7A35019C36D64 CRC64;

Query Match 30.4%; Score 647.5; DB 4; Length 353;
Best Local Similarity 39.7%; Pred. No. 4.7e-51;
Matches 147; Conservative 53; Mismatches 115; Indels 55; Gaps 11;

QY 40 GVIIPEKFKWPPSTPRA---YNNREQEKLNRWYNPILNRVANQTGELATSPNTSHLSYC 96
DB 19 GCLLFVRKTAAGDPTAHQPFW-----APTPRHRSRC 51
QY 97 EPDSTVMTAVTDNNLPDRFKDFLLYLCRCNYSLLIDQPKKCAKKPFLLLAIKSLIPHEA 156
DB 52 PPNHTVSSASL---SLPSRHLFLTYRHCNFSILL-EPSCGSKDTFLLAIAKSPGHVE 107
QY 157 RQAIRESWGRETNVG-----NQTVVVRVFL---GKTPPEDNHPDLSMDMLKFESDKHQDIL 209
DB 108 RRAAIRSTWGR---VGGWARGROLKLVFLGVAGSAPP-----AQLLAYESREFDDIL 157
QY 210 MNRYRTFFNLISLKEVLFRLWVSTSCPDAEFVKGDGDDVFVNTHILNLSLSKSKAKD 269
DB 158 QWDETEDEFFNLTKELHLQRWVVAACPAHFMLKGGDDVFVHPNVLEFDGWN--DPAQD 215
QY 270 LFGDVTNAGPRHKKLYIPEVFTGV-YPPYAGGGGFLYSGPALLRLYSATSRRVHL 328
DB 216 LLVGDTVIRQALPNRNTKVYFIPPSMYRATHYPPYAGGGGVYMSRATVRRLOAIMEDAEL 275
QY 329 YPIDDVYTGMCLOKLGVLPEKHKGFRTFDEEK-NKNKICSYIDLMLVHSRKPQEMIDIW 387
DB 276 FPIDDVFGVGMCLRLGLSPMHAGFKTFGIRRLDPLDPCLYRGLLLVHRLSPLEMTWM 335
QY 388 SLOQSPNLKC 397
DB 336 ALVTDEGLKC 345

RESULT 15
Q9COJ1 PRELIMINARY; PRT; 378 AA.
AC Q9COJ1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Beta-1.3-N-acetylglucosaminyltransferase B6n-T4.
GN B6NT-4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shiraishi N., Natsume A., Toyayachi A., Endo T., Akashima T.,
RA Yamada Y., Imai N., Nakagawa S., Koizumi S., Sekine S., Narimatsu H.,
RA Sasaki K.;
RT *Identification and Characterization of Three Novel beta1,3-N-
RT Acetylglucosaminyltransferases Structurally Related to the beta1,3-
RT Galactosyltransferase Family.*;
RL J. Biol. Chem. 276:3498-3507(2001).
DR EMBL; AB049586; BAB21532.1;
DR Genew; HGNC:15683; B3GNT4;
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 378 AA; 42310 MW; E8E88665A36057E CRC64;

Query Match 30.4%; Score 647.5; DB 4; Length 378;
Best Local Similarity 39.7%; Pred. No. 5.2e-51;
Matches 147; Conservative 53; Mismatches 115; Indels 55; Gaps 11;

QY 40 GVIIPEKFKWPPSTPRA---YNNREQEKLNRWYNPILNRVANQTGELATSPNTSHLSYC 96
DB 44 GCLLFVRKTAAGDPTAHQPFW-----APTPRHRSRC 76
QY 97 EPDSTVMTAVTDNNLPDRFKDFLLYLCRCNYSLLIDQPKKCAKKPFLLLAIKSLIPHEA 156
DB 77 PPNHTVSSASL---SLPSRHLFLTYRHCNFSILL-EPSCGSKDTFLLAIAKSPGHVE 132
QY 157 RQAIRESWGRETNVG-----NQTVVVRVFL---GKTPPEDNHPDLSMDMLKFESDKHQDIL 209
DB 133 RRAAIRSTWGR---VGGWARGROLKLVFLGVAGSAPP-----AQLLAYESREFDDIL 182

```

```

Qy 210 MMYRDTFFNLSKEVLFLEWVSTSGPDAEFVFKGDDVFVNTHTILNYLNSLSKSKAKD 269
Db 183 QWDFTEDFNLILKELHLQRWVVAACQAHFMUKGDDVFVHPNVLEFDGW--DPAOD 240
Qy 270 LFIGDVHINAGPHRDKKLYYIPEVFTGV-YPPYAGGGGGLYSGPALLRLYSATSRVHL 328
Db 241 LLVGDVIRQALPNRNTKVYFIPPSMYRATHYPPYAGGGGYVMSRATVRRLOAIMEDAE 300
Qy 329 YPIDDVYTGNCLOKGLGLVPEKHKGFRTDIEEK-NKKNICSYIDLMLVHSRKPQEMIDIW 387
Db 301 FPIDDVFGVGLRRLGLSPMHAGFKTFGIRRLDPLDPCLYRGLLVHRLSPLEMTW 360
Qy 388 SQLQSPNLKC 397
Db 361 ALVTDEGLKC 370

```

Search completed: October 20, 2003, 23:04:02
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: October 20, 2003, 22:49:49 ; Search time 38.5 Seconds

(without alignments)
1636.739 Million cell updates/sec

Title: US-09-804-357b-14

Perfect score: 2123

Sequence: 1 MSVGRRRIKLLGILMANVF.....RKPEMIDWSQLQSAHLAC 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2123	100.0	397	19	AAW80212 Human cardiac and
2	2123	100.0	397	21	AAAY84683 Amino acid sequenc
3	2123	100.0	397	22	AAAG66118 Human LiG46 polype
4	2123	100.0	397	23	AAAE29093 Human LiG46 protei
5	2123	100.0	397	24	ABAB2649 Human beta3gnt pol
6	2105	98.2	413	22	ABAB49748 Human beta 1,3-N-a
7	2022	95.2	397	21	AAAY69698 Human brainiac pro
8	2007	94.5	373	22	AAAB93536 Human protein sequ
9	1872	88.2	397	21	AAAY84641 Amino acid sequenc

10	1872	88.2	397	21	AAAY79953	Murine brainiac pr
11	1872	88.2	397	21	AAAY69697	Murine Brainiac pr
12	1872	88.2	397	22	AAAG66115	Murine LiG46 polyp
13	1872	88.2	397	23	AAAE29092	Murine LiG46 prote
14	1866	87.9	397	24	ABAB2649	Murine beta3gnt po
15	1823	85.9	397	21	AAAY79954	Human brainiac pro
16	1743	82.1	367	22	AAAG66116	Murine LiG46 matur
17	1068	50.3	197	23	AAAE15936	Human 7023630/124-
18	667	31.4	401	23	AAE22148	Human TRNFR-10 pro
19	642.5	30.3	364	23	ABU65191	Human NOV106a prot
20	641.5	30.2	334	22	ABG20251	Novel human diagno
21	634.5	29.9	415	23	ABG30984	Human galactosyltr
22	632	29.8	352	20	AAAY06462	Human Brainiac-3.
23	632	29.8	352	24	ABR47731	Human secreted pro
24	632	29.8	353	20	AAAY36224	Human secreted pro
25	628.5	29.6	378	21	AAAB24033	Human PRO4344 prot
26	628.5	29.6	378	22	AAAB49751	Human beta 1,3-N-a
27	628.5	29.6	378	23	ABG34047	Human pro peptide
28	616.5	29.0	378	21	AAAB07435	The beta-1,3-galac
29	616.5	29.0	378	23	ABG32941	Human galactosyltr
30	616.5	29.0	378	23	AAE24688	Human znssp6 prote
31	614.5	28.9	384	23	ABG32374	Novel human enzyme
32	602	28.4	402	21	ABG24035	Human PRO4397 prot
33	602	28.4	402	23	AAE22142	Human TRNFR-4 prot
34	602	28.4	402	23	ABG34046	Human pro peptide
35	601	28.3	372	19	AAW64558	Human epidermoid c
36	601	28.3	372	22	AAU29167	Human p80 polypept
37	601	28.3	372	22	AAAB49749	Human beta 1,3-N-a
38	601	28.3	372	23	ABAB09716	Amino acid sequenc
39	601	28.3	372	23	AAU11272	Human beta1,3-N-ac
40	601	28.3	372	24	ABU71255	Human PRO1266 prot
41	601	28.3	372	24	ABU65712	Human secreted/tra
42	601	28.3	372	24	ABU66045	Novel human secret
43	601	28.3	372	24	ABU67549	Human secreted/tra
44	601	28.3	372	24	ABR01801	Human cancer-relat
45	601	28.3	372	24	ABU65407	human p80 polypept

ALIGNMENTS

RESULT 1

AAW80212 ID AAW80212 standard; Protein; 397 AA.

XX AAW80212;

XX AC

XX AC

DI 18-JAN-1999 (first entry)

XX Human cardiac and pancreatic protein (CAPP).

DE Cardiac and pancreatic protein; CAPP; muscle-derived growth factor;

XX human; pancreatitis; myocardial infarction; cardiomyopathy.

KW Homo sapiens.

OS Homo sapiens.

XX Key

XX Location/Qualifiers

XX 1..32

XX /label= Sig_peptide

XX 33..397

XX /label= Mat_protein

XX 1..11

XX /note= "epitope-bearing peptide (Claim 11)"

XX 29..72

XX /note= "epitope-bearing peptide (Claim 11)"

XX 78..89

XX /note= "epitope-bearing peptide (Claim 11)"

XX 94..105

XX /note= "epitope-bearing peptide (Claim 11)"

XX 110..119

XX /note= "epitope-bearing peptide (Claim 11)"

XX 124..142

XX /note= "epitope-bearing peptide (Claim 11)"

XX /note= "epitope-bearing peptide (Claim 11)"

```

FT Peptide 151..176
FT /note= "epitope-bearing peptide (Claim 11)"
FT Peptide 184..218
FT /note= "epitope-bearing peptide (Claim 11)"
FT Peptide 232..251
FT /note= "epitope-bearing peptide (Claim 11)"
FT Peptide 262..272
FT /note= "epitope-bearing peptide (Claim 11)"
FT Peptide 280..290
FT /note= "epitope-bearing peptide (Claim 11)"
FT Peptide 346..368
FT /note= "epitope-bearing peptide (Claim 11)"
FT Peptide 376..385
FT /note= "epitope-bearing peptide (Claim 11)"
XX
XX WO9844112-A1.
XX
XX 08-OCT-1998.
XX
XX 27-MAR-1998; 98WO-US06022.
XX
XX 28-MAR-1997; 97US-0042855.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Soppet DR;
XX
XX WPI: 1998-557111/47.
XX
XX N-PSDB; AAV66367.
XX
XX New isolated cardiac and pancreatic protein - used to develop
XX products for the diagnosis and treatment of e.g. pancreatitis or
XX abnormal hypertrophy of the heart
XX
XX Claim 1: Fig 1: 113pp; English.
XX
XX This is the amino acid sequence of novel cardiac and pancreatic
XX protein (CAPP), a member of the muscle-derived growth factor
XX superfamily, as deduced from the nucleotide sequence (see AAV66367)
XX of a cDNA clone discovered in a cDNA library derived from activated
XX T cells. The gene is abundant in adult heart and pancreas. CAPP
XX isolated nucleic acid molecules and polypeptides are provided, as
XX are vectors, host cells and recombinant methods for producing the
XX same. The invention further relates to screening methods for
XX identifying agonists and antagonists of CAPP activity, as well as
XX diagnostic methods for detecting the presence of activated T-cells,
XX and mature heart, pancreas and placental tissues and cells. The
XX CAPP polypeptides can modulate the differentiation and proliferation
XX of cells and tissue, both in vivo and ex vivo. The products can be
XX used in the diagnosis and treatment of pancreatitis and conditions
XX that cause abnormal hypertrophy of the heart, such as hypertension,
XX myocardial infarction, valve disease and cardiomyopathy. The
XX products can also be used in detection and cell culturing.
XX
XX Sequence 397 AA;
XX
XX Query Match 100.0%; Score 2123; DB 19; Length 397;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-218;
XX Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MSVGRRIKILGLILMANVFIYIMEVSKSSQEKNGKEVIIPKPKFWKISTPPPAYWN 60
XX 1 MSVGRRIKILGLILMANVFIYIMEVSKSSQEKNGKEVIIPKPKFWKISTPPPAYWN 60
XX
XX 61 REQELNROYNPILSLMTNOTGAGRLSNLSHLYNCEPDLRVTSVVTGFNNLPDRKDEL 120
XX
XX 61 REQELNROYNPILSLMTNOTGAGRLSNLSHLYNCEPDLRVTSVVTGFNNLPDRKDEL 120
XX
XX 121 LYLRCRNYSLIDQDKCAKKPFLLAIKSLTFFHARRQAIRESWGQESNAGNQTVRVVF 180
XX
XX 121 LYLRCRNYSLIDQDKCAKKPFLLAIKSLTFFHARRQAIRESWGQESNAGNQTVRVVF 180
XX
XX 181 LLGOTPPEDNHPDLSMLKFESEKHQDILMNNYRDTFFNLSLKEVFLRWVSTSCPDTEF 240
XX

```



```
Query Match      100.0%; Score 2123; DB 21; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.4e-218;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVIIPKEKFWKISTPPEAYWN 60
DB 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVIIPKEKFWKISTPPEAYWN 60
QY 61 REQELNRYNPILSMLTNOTGAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDFL 120
DB 61 REQELNRYNPILSMLTNOTGAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDFL 120
QY 121 LYLCRNYSLIDQPKCAKPPFLLLAISKLTSPHFARRQAIRESWGQESNAGNQTIVRVF 180
DB 121 LYLCRNYSLIDQPKCAKPPFLLLAISKLTSPHFARRQAIRESWGQESNAGNQTIVRVF 180
QY 181 LLGQTPPEDNHPDLSMLKFESEKHQDILMNNYRTDFFNLSKEVFLFRWVSTSCPDTEF 240
DB 181 LLGQTPPEDNHPDLSMLKFESEKHQDILMNNYRTDFFNLSKEVFLFRWVSTSCPDTEF 240
QY 241 VFKGDDVFNTHILNLSKTKAKDLFIGDVIHNAIPHRRDKKLYIPEVYVSGLY 300
DB 241 VFKGDDVFNTHILNLSKTKAKDLFIGDVIHNAIPHRRDKKLYIPEVYVSGLY 300
QY 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVTGMCCLKLGLVPEKHKGFRTDIEE 360
DB 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVTGMCCLKLGLVPEKHKGFRTDIEE 360
QY 361 KNKNKICSYVDLMLVHSRKPQEMIDWSQLSAHLKC 397
DB 361 KNKNKICSYVDLMLVHSRKPQEMIDWSQLSAHLKC 397

RESULT 3
AAG66118
ID AAG66118 standard; Protein; 397 AA.
AC AAG66118;
XX 13-MAR-2002 (first entry)
XX Human LIG46 polypeptide.
XX Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Stral3;
XX anorectic; anabolic; antisense therapy; human.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Peptide 1..32
XX Protein /note= "signal peptide"
XX /note= "mature protein"
XX US2001024808-A1.
XX 27-SEP-2001.
XX 12-MAR-2001; 2001US-0804357.
XX 29-OCT-1998; 98US-106378P.
XX 19-NOV-1998; 98US-0195896.
XX 10-SEP-1998; 98US-0150857.
XX (MILL-) MILLENNIUM PHARM INC.
XX White D, Zhou J, Tartaglia LA;
XX WPI: 2001-624963/72.
XX N-PSDB; AA167869.
XX An isolated polypeptide useful for treating a weight disorder and for
```

```
PT screening compounds that may inhibit protein activity comprises a
PT leptin-induced protein
XX Example 2; Fig 8; 46pp; English.
XX The invention relates to genes whose expression are induced by leptin.
XX LIG46 and LIG56 are novel leptin induced genes (LIG), while four other
XX genes such as Tgtp, LRG-47, RC10-II and Stral3 have been previously
XX identified. The leptin induced proteins can be expressed by standard
XX recombinant methodology. The proteins and encoding polynucleotides may
XX be used in screening assays to identify compounds that may bind to it.
XX Administering a molecule, e.g., an antisense molecule, which reduces
XX expression of activity of protein selected from LIG46, LIG56, Tgtp,
XX LRP-47, RC10-II, and Stral3 may be used to treat a weight disorder. The
XX method may further comprise administering leptin. The present sequence
XX represents a human LIG46 protein.
SQ Sequence 397 AA;
Query Match      100.0%; Score 2123; DB 22; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.4e-218;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVIIPKEKFWKISTPPEAYWN 60
DB 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVIIPKEKFWKISTPPEAYWN 60
QY 61 REQELNRYNPILSMLTNOTGAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDFL 120
DB 61 REQELNRYNPILSMLTNOTGAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDFL 120
QY 121 LYLCRNYSLIDQPKCAKPPFLLLAISKLTSPHFARRQAIRESWGQESNAGNQTIVRVF 180
DB 121 LYLCRNYSLIDQPKCAKPPFLLLAISKLTSPHFARRQAIRESWGQESNAGNQTIVRVF 180
QY 181 LLGQTPPEDNHPDLSMLKFESEKHQDILMNNYRTDFFNLSKEVFLFRWVSTSCPDTEF 240
DB 181 LLGQTPPEDNHPDLSMLKFESEKHQDILMNNYRTDFFNLSKEVFLFRWVSTSCPDTEF 240
QY 241 VFKGDDVFNTHILNLSKTKAKDLFIGDVIHNAIPHRRDKKLYIPEVYVSGLY 300
DB 241 VFKGDDVFNTHILNLSKTKAKDLFIGDVIHNAIPHRRDKKLYIPEVYVSGLY 300
QY 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVTGMCCLKLGLVPEKHKGFRTDIEE 360
DB 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVTGMCCLKLGLVPEKHKGFRTDIEE 360
QY 361 KNKNKICSYVDLMLVHSRKPQEMIDWSQLSAHLKC 397
DB 361 KNKNKICSYVDLMLVHSRKPQEMIDWSQLSAHLKC 397

RESULT 4
AAE29093
ID AAE29093 standard; Protein; 397 AA.
XX AAE29093;
XX AAE29093;
XX 27-JAN-2003 (first entry)
XX Human LIG46 protein.
XX LIG46; leptin; low body weight; chromosomal mapping; tissue typing;
XX forensic biology; transgenic; gene therapy; antianorectic; human;
XX chromosome 2.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Peptide 1..32
XX Protein /label= signal_peptide
XX /note= "Human mature LIG46 protein"
```

```
XX WO200274905-A2.
XX 26-SEP-2002.
XX 20-NOV-2001; 2001WO-US43345.
XX 21-NOV-2000; 2000US-0717778.
XX (MILL-) MILLENNIUM PHARM INC.
XX White DW, Zhou J, Tartaglia LA, Stricker-krongrad A, Clausen H;
XX WPI: 2002-759886/82.
XX N-PSDB; AAD46673.
XX New isolated nucleic acid molecules, designated as leptin-induced genes
XX 46 (LIG46), useful for treating a subject having a disorder
XX characterized by undesirable level of LIG46 expression or activity,
XX such as low body weight.
XX Claim 22; Fig 6; 90pp; English.
XX The invention relates to LIG46, a gene whose expression is induced by
XX leptin. LIG46 DNA and protein are useful in treating a subject having
XX a disorder characterized by undesirable level of LIG46 expression or
XX activity, such as low body weight. They are also useful in a screening
XX assay, chromosomal mapping, tissue typing and forensic biology. The
XX probes based on the LIG46 nucleotide sequence are useful for detecting
XX transcripts or genomic sequences encoding the same or related proteins.
XX The LIG46 polypeptides are useful as immunogens for raising anti-LIG46
XX antibodies. The host cells are useful for producing non-human transgenic
XX animals. LIG46 DNA is used in gene therapy. The present sequence is
XX human LIG46 protein. LIG46 gene is located at chromosome 2.
XX
XX Sequence 397 AA:
Query Match 100.0%; Score 2123; DB 23; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.4e-218;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVFIIPKEFKWISTPPEAYWN 60
DB 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVFIIPKEFKWISTPPEAYWN 60
QY 61 REQEKLNRQYNPILSMLTNOTGEAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRKDFL 120
DB 51 REQEKLNRQYNPILSMLTNOTGEAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRKDFL 120
QY 121 LYLRCRNYSLIDQPKCAKKPFLLLAISKLTPIHARRQAIRESWGQESNAGNQTIVRVF 180
DB 121 LYLRCRNYSLIDQPKCAKKPFLLLAISKLTPIHARRQAIRESWGQESNAGNQTIVRVF 180
QY 181 LLGQTPPEDNHPDLSMDLKFESEKHQDILMWNRTDFNLSLKEVFLRWVSTSCPDTEF 240
DB 181 LLGQTPPEDNHPDLSMDLKFESEKHQDILMWNRTDFNLSLKEVFLRWVSTSCPDTEF 240
QY 241 VFKGDDDFVNTHTLNLNLSKTKAKOLFIDGVIHNAGPHRDKKLYIPEVYISGLY 300
DB 241 VFKGDDDFVNTHTLNLNLSKTKAKOLFIDGVIHNAGPHRDKKLYIPEVYISGLY 300
QY 301 PPYAGGGFLYSGHLARLTHITDQVHLYPIDDVYTCMLQKLGVLPEKHKGFRTEIDEE 360
DB 301 PPYAGGGFLYSGHLARLTHITDQVHLYPIDDVYTCMLQKLGVLPEKHKGFRTEIDEE 360
QY 361 KNKNKNCISYDMLVHRSRKPQEMDIDWSQSAHLKC 397
DB 361 KNKNKNCISYDMLVHRSRKPQEMDIDWSQSAHLKC 397
RESULT 5
ABB82649
ID ABB82649 standard; Protein; 397 AA.
```

```
XX ABB82649;
XX 19-FFB-2003 (first entry)
XX Human beta3Gnt polypeptide sequence.
XX Beta3Gnt; transgenic; beta-1,3-N-acetylglucosaminyltransferase;
XX disease system; human; enzyme.
XX Homo sapiens.
XX WO200279413-A2.
XX 10-OCT-2002.
XX 29-MAR-2002; 2002WO-US09645.
XX 29-MAR-2001; 2001US-280706P.
XX 28-MAR-2002; 2002US-0280706.
XX (DELT-) DELTAGEN INC.
XX Leviten MW, Phillips R;
XX WPI: 2003-067437/06.
XX N-PSDB; ABV75082.
XX New transgenic mouse comprising a disruption in a beta3Gnt gene, as in
XX vivo model to study various disease states or conditions in which
XX beta3Gnt may be implicated or involved, such as abnormal cell growth,
XX cancer and metastasis.
XX Disclosure; Fig 4; 55pp; English.
XX The invention relates to a transgenic mouse comprising a disruption in a
XX beta3Gnt (beta-1,3-N-acetylglucosaminyltransferase) gene, where there is
XX no native expression of beta3Gnt gene. The transgenic mice may be used
XX as in vivo model to study various disease states or conditions in which
XX beta3Gnt may be implicated or involved, such as abnormal cell growth,
XX cancer and metastasis, and to evaluate various treatments or to identify
XX agents for treating disease states or conditions, such as anxiety or
XX depression. Animal-based disease systems may be used to identify
XX compounds capable of ameliorating disease symptoms, as test substrates
XX for the identification of drugs, pharmaceuticals, therapies and
XX interventions that may be effective in treating a disease or other
XX phenotypic characteristic of the animal. The present sequence represents
XX a human beta3Gnt polypeptide sequence.
XX
XX Sequence 397 AA:
Query Match 100.0%; Score 2123; DB 24; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.4e-218;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVFIIPKEFKWISTPPEAYWN 60
DB 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVFIIPKEFKWISTPPEAYWN 60
QY 61 REQEKLNRQYNPILSMLTNOTGEAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRKDFL 120
DB 61 REQEKLNRQYNPILSMLTNOTGEAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRKDFL 120
QY 121 LYLRCRNYSLIDQPKCAKKPFLLLAISKLTPIHARRQAIRESWGQESNAGNQTIVRVF 180
DB 121 LYLRCRNYSLIDQPKCAKKPFLLLAISKLTPIHARRQAIRESWGQESNAGNQTIVRVF 180
QY 181 LLGQTPPEDNHPDLSMDLKFESEKHQDILMWNRTDFNLSLKEVFLRWVSTSCPDTEF 240
DB 181 LLGQTPPEDNHPDLSMDLKFESEKHQDILMWNRTDFNLSLKEVFLRWVSTSCPDTEF 240
QY 241 VFKGDDDFVNTHTLNLNLSKTKAKOLFIDGVIHNAGPHRDKKLYIPEVYISGLY 300
DB 241 VFKGDDDFVNTHTLNLNLSKTKAKOLFIDGVIHNAGPHRDKKLYIPEVYISGLY 300
```

DD 241 VFKGDDVFNTHILNLSLTKAKDLFIGDVHNNAGPHRDKLKYIPEVYVSGLY 300
 QY 301 PPVAGGGFLYSGHLALRLYHITDQVHLYPIDVYTGMLQKLGVLVPEKHKGFRFTD:EE 360
 DB 301 PPVAGGGFLYSGHLALRLYHITDQVHLYPIDVYTGMLQKLGVLVPEKHKGFRFTD:EE 360
 QY 361 KKNKNICSVDMLVHRSRKPQEMIDTWSQLQSAHLKC 397
 DB 361 KKNKNICSVDMLVHRSRKPQEMIDTWSQLQSAHLKC 397

RESULT 6
 AAB49748
 ID AAB49748 standard; protein: 413 AA.
 XX AC AAB49748;
 XX DT 17-APR-2001 (first entry)
 XX DE Human beta 1,3-N-acetylglucosamine transferase protein G3.
 XX KW Sugar chain synthesis agent; beta 1,3-N-acetylglucosamine transferase;
 XX KW inflammation; cancer; metastasis; human.
 XX OS Homo sapiens.
 XX PN WO200100848-A1.
 XX PD 04-JAN-2001.
 XX PF 29-JUN-2000; 2000WO-JP04304.
 XX PR 29-JUN-1999; 99JP-0183437.
 XX PR 16-MAR-2000; 2000JP-0074757.
 XX PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX PI Sasaki K, Shiraishi N, Natsume A, Yamada Y, Nakagawa S, Sekine S;
 XX WPI; 2001-102895/11.
 XX DR N-PSDB; AAF29255.
 XX PT New polypeptide having beta1,3-N-acetylglucosamine transferase activity
 PT for diagnosis of inflammation, cancer and cancer metastasis,
 PT development of remedies, and for producing glycoconjugates -
 XX PS Claim 1; Page 147-150; 195pp; Japanese.
 XX CC This invention relates to a sugar chain synthesizing agent that contains
 CC a polypeptide as the active ingredient, where the polypeptide has beta
 CC 1,3-N-acetylglucosamine transferase activity. The polypeptides, DNA
 CC sequences encoding them and antibodies directed against the proteins are
 CC useful in the diagnosis of inflammation, cancer and its metastasis,
 CC development of remedies, and for producing sugar chains and
 CC glycoconjugates. Sequences AAB49748 - AAB49751 represent polypeptides of
 CC the invention, having beta 1,3-N-acetylglucosamine transferase activity,
 CC and polynucleotides AAF29255 - AAF29258 represent cDNA encoding them. PCR
 CC primers used in the isolation and characterisation of the cDNA sequences
 CC are represented by sequences ANF29259 - AAF29290.
 XX SQ Sequence 413 AA;
 Query Watch 99.2%; Score 2105; DB 22; Length 413;
 Best Local Similarity 96.1%; Pred. No. 1.3e-216;
 Matches 397; Conservative 0; Mismatches 0; Indels 16; Gaps 1;
 QY 1 MSVGRRRKLLGILM-----ANVFIFYFIMEVSKSSQKNGKGEVIIP 44
 DB 1 MSVGRRRKLLGILMANNVFIFYFIMEVSKSSANVFIFYFIMEVSKSSQKNGKGEVIIP 60
 QY 45 KEKFWKISTPPEAYNREQELNRYNPILSMLTNQTGEAGRLSNISHLNYPEDLRVTS 104
 DB 61 KEKFWKISTPPEAYNREQELNRYNPILSMLTNQTGEAGRLSNISHLNYPEDLRVTS 120

QY 105 VVTGFNNLPDRFKDFFLLYLRRCRNYSLIDQPKCAKKPFLLLAIKSLTSPHFARROAIRES 164
 DB 121 VVTGFNNLPDRFKDFFLLYLRRCRNYSLIDQPKCAKKPFLLLAIKSLTSPHFARROAIRES 180
 QY 165 WGESNAGNQTIVRVFLLGQTTPPEDNHPDLSMDLKFESEKHQDILMWNKYDFTFNLSLKE 224
 DB 181 WGESNAGNQTIVRVFLLGQTTPPEDNHPDLSMDLKFESEKHQDILMWNKYDFTFNLSLKE 240
 QY 225 VLELRWYSTSCPDTEFVFVKGDDVFNTHILNLSLTKAKDLFIGDVHNNAGPHRD 284
 DB 241 VLELRWYSTSCPDTEFVFVKGDDVFNTHILNLSLTKAKDLFIGDVHNNAGPHRD 300
 QY 285 KKLKYIPEVYVSGLYPPYAGGGFLYSGHLALRLYHITDQVHLYPIDVYTGMLQKLG 344
 DB 301 KKLKYIPEVYVSGLYPPYAGGGFLYSGHLALRLYHITDQVHLYPIDVYTGMLQKLG 360
 QY 345 LVPEKHKGFRFTDIEEKNKNKNICSVDMLVHRSRKPQEMIDTWSQLQSAHLKC 397
 DB 361 LVPEKHKGFRFTDIEEKNKNKNICSVDMLVHRSRKPQEMIDTWSQLQSAHLKC 413

RESULT 7
 AAY69698
 ID AAY69698 standard; Protein; 397 AA.
 XX AC AAY69698;
 XX DT 08-MAY-2000 (first entry)
 XX DE Human Brainiac protein.
 XX KW Brainiac; human; mammalian; expressed sequence tag; EST;
 KW Drosophila melanogaster; Egghead; Notch; epithelial cell adhesion;
 KW Fringe family; cell viability; growth regulation; cell fate; cancer;
 KW psoriasis; skin lesion; nervous system disorder; developmental syndrome;
 KW transgenic animal; drug screening.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Misc-difference 104 /note= "Encoded by NNNNNN"
 FT CA2255109-A1.
 XX PD 17-JUN-1999.
 XX PF 17-DEC-1998; 98CA-2255109.
 XX PR 17-DEC-1997; 97CA-2225126.
 XX PA (HSCR-) HSC RES & DEV LP.
 XX PI Egan SE;
 XX WPI; 2000-148082/14.
 XX DR N-PSDB; AAB67186.
 XX PT New nucleic acids encoding a murine and human Brainiac protein, useful
 PT for detecting somatic or germline DNA lesions which are responsible for
 PT developmental syndromes or diseases including cancer -
 XX PS Claim 13; Page31; 40pp; English.
 XX CC This sequence represents human Brainiac protein. The complete murine
 CC Brainiac cDNA sequence (AA287185) was used to screen an EST (expressed
 CC sequence tag) database to identify human Brainiac cDNA clones,
 CC which were assembled to form the human Brainiac cDNA of the present
 CC invention. Human Brainiac has significant similarity to Drosophila
 CC melanogaster Brainiac, Drosophila Brainiac and Egghead proteins regulate
 CC adhesion between epithelial cells, this activity being dependent on the
 CC presence of Notch. Drosophila Brainiac is a secreted protein, and has

CC sequence similarities with the Drosophila Fringe proteins. Brainiac/
CC Egghead-mediated epithelial cell adhesion is required for cell
CC viability, cell growth regulation and cell fate specification. Wild-type
CC or mutant forms of mammalian Brainiac proteins may therefore be used to
CC alter epithelial cell adhesion in a mammal. Mammalian Brainiac proteins,
CC active fragments analogues, and nucleic acids may be used to treat
CC diseases such as cancer, psoriasis and other skin lesions, and nervous
CC system disorders. Mammalian Brainiac nucleic acids may also be used to
CC detect somatic or germline DNA lesions which are responsible for
CC developmental syndromes or diseases including cancer. The mammalian
CC Brainiac proteins and fragments or its analogues are useful as antigens
CC in immunoassays including enzyme-linked immunosorbent assays (ELISA),
CC radioimmunoassays (RIA) and other non-enzyme linked antibody binding
CC assays. Non-human transgenic animals comprising nucleotide sequences
CC encoding human Brainiac protein (AA96969) can be used as animal models
CC for the study of mammalian Brainiac gene function, for the screening of
CC candidate compounds and for the evaluation of potential therapeutic
CC interventions.

XX Sequence 397 AA;

Query Match 95.2%; Score 2022; DB 21; Length 397;
Best Local Similarity 97.7%; Pred. No. 9.5e-208;
Matches 379; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 10 LIGILMANVYFIYINVEVSKSSQKNGKGEVLIIPKEFWKISTPPEAYWNRQEKLNQ 69
DB 10 LIGILMANVYFIYINVEVSKSSQKNGKGEVLIIPKEFWKISTPPEAYWNRQEKLNQ 69
QY 70 YNPILSMLTNGEAGRLSNISHLNCEPDLRVTSVVTGFNNLPDRFDFLLYLCRNY 129
DB 70 YNPILSMLTNGEAGRLSNISHLNCEPDLRVTSVVTGFNNLPDRFDFLLYLCRNY 129
QY 130 LLTDQPKCAKPFLLAIAKSLTPHARRQAIRESWGQESNAGNQTIVRVFLLGOTPPED 189
DB 130 LLTDQPKCAKPFLLAIAKSLTPHARRQAIRESWGQESNAGNQTIVRVFLLGOTPPED 189
QY 190 NHPDLSMLKFESKQHDILMNNYRDTFFNLSLKEVLFRLWYSTSCPTDFVFKGDDVDF 249
DB 190 NHPDLSMLKFESKQHDILMNNYRDTFFNLSLKEVLFRLWYSTSCPTDFVFKGDDVDF 249
QY 250 VNTTHILNLSLKTAKDLFIGDVHINAGPHRDKKLYIPEVYVSGLYPPYAGGGGF 309
DB 250 VNTTHILNLSLKTAKDLFIGDVHINAGPHRDKKLYIPEVYVSGLYPPYAGGGGF 309
QY 310 LYSGLHALRLYHITDOVHLYPIDDVYTGMLQKLGVLPEKHKGFRTDIEKNKNNICS 369
DB 310 LYSGLHALRLYHITDOVHLYPIDDVYTGMLQKLGVLPEKHKGFRTDIEKNKNNICS 369
QY 370 VDLMLVHSRKPQEMIDVWSQLOSAHLKC 397
DB 370 VDLMLVHSRKPQEMIDVWSQLOSAHLKC 397

RESULT 8
AAB93536

ID AAB93536 standard; Protein; 373 AA.

XX AAB93536;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12898.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

KW Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX

PS Claim 8; SEQ ID 12898; 25377pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination
CC of the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

SQ Sequence 373 AA;

Query Match 94.5%; Score 2007; DB 22; Length 373;
Best Local Similarity 100.0%; Pred. No. 3.5e-206;
Matches 373; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 MEVSKSSQKNGKGEVLIIPKEFWKISTPPEAYWNRQEKLNQYNPILSMLTNGEA 84
DB 1 MEVSKSSQKNGKGEVLIIPKEFWKISTPPEAYWNRQEKLNQYNPILSMLTNGEA 60
QY 85 GRLSNISHLNCEPDLRVTSVVTGFNNLPDRFDFLLYLCRNYSLIIDQPKCAKPF 144
DB 61 GRLSNISHLNCEPDLRVTSVVTGFNNLPDRFDFLLYLCRNYSLIIDQPKCAKPF 120
QY 145 LLAIKSLTPHARRQAIRESWGQESNAGNQTIVRVFLLGOTPPEDNHPDLSOMLKFESK 204
DB 121 LLAIKSLTPHARRQAIRESWGQESNAGNQTIVRVFLLGOTPPEDNHPDLSOMLKFESK 180
QY 205 HQDILMNNYRDTFFNLSLKEVLFRLWYSTSCPTDFVFKGDDVDFVNTTHILNLSL 264
DB 181 HQDILMNNYRDTFFNLSLKEVLFRLWYSTSCPTDFVFKGDDVDFVNTTHILNLSL 240
QY 265 TKADLFIGDVHINAGPHRDKKLYIPEVYVSGLYPPYAGGGGFYSGLHALRLYHITD 324
DB 241 TKADLFIGDVHINAGPHRDKKLYIPEVYVSGLYPPYAGGGGFYSGLHALRLYHITD 300
QY 325 QVHLYPIDDVYTGMLQKLGVLPEKHKGFRTDIEKNKNNICSYYVLDMLVHSRKPQEMI 384
DB 301 QVHLYPIDDVYTGMLQKLGVLPEKHKGFRTDIEKNKNNICSYYVLDMLVHSRKPQEMI 360

QY 385 DIWSQLOSAHLKC 397
 DB 361 DIWSQLOSAHLKC 373

RESULT 9
 AAY84641
 ID AAY84641 standard; Protein: 397 AA.
 AC AAY84641;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of a murine LIG46 polypeptide.
 XX
 KW Leptin; LIG46; body weight; leptin inducible gene; obesity; cachexia;
 KW LIG56; Tgtp; LRG-47; RC10-II; Stral3.
 XX
 OS Mus sp.
 XX
 PN W0200015826-A2.
 XX
 PD 23-MAR-2000.
 XX
 PF 10-SEP-1999; 99WO-US20722.
 XX
 PR 10-SEP-1998; 98US-0150857.
 PR 29-OCT-1998; 98US-0106378.
 PR 19-NOV-1998; 98US-0195896.
 PR 15-APR-1999; 99US-0292228.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI White D, Zhou J, Tartaglia LA;
 XX
 DR WPI: 2000-271461/23.
 DR N-PSDB; AAA12714.
 XX
 PT Method for determining compounds which modulate body weight and can be
 PT used to treat e.g. obesity comprises measuring the activity of leptin
 PT inducible genes .
 XX
 PS Claim 23; Fig 1A-B; 123pp; English.
 XX
 CC The present sequence represents a leptin induced LIG46 polypeptide. The
 CC specification describes a method for determining whether a compound can
 CC be used to modulate body weight by measuring the activity of leptin
 CC inducible genes, such as LIG46. The method can be used to specifically
 CC detect LIG46 and LIG56 nucleic acid molecules relative to other nucleic
 CC acid molecules encoding galactosyltransferases or GTP-binding proteins.
 CC They can also be used in diagnostic assays to identify the presence or
 CC absence of a genetic lesion or mutation characterized by aberrant
 CC modification or mis-regulation of the genes or aberrant post-
 CC translational modification of the proteins. LIG46 and LIG56 proteins
 CC and nucleic acid molecules can be used to treat obesity or cachexia.
 CC LIG46 and LIG56 antagonists are used to treat obesity and LIG46 and
 CC LIG56 agonists are used to treat low body weight. The leptin inducible
 CC genes i.e. LIG46, LIG56, Tgtp, LRG-47, RC10-II and Stral3 can be used
 CC to determine whether a compound modulates body weight and can then be
 CC used to treat obesity or cachexia or low body weight.
 XX
 SQ Sequence 397 AA;

Query Match 88.28; Score 1872; DB 21; Length 397;
 Best Local Similarity 87.48; Pred. No. 1.2e-191;
 Matches 347; Conservative 13; Mismatches 31; Indels 0; Gaps 0;

QY 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKGEVIIPKEKFKWIKSTPPAYWN 60
 DB 1 MSVGRRIKLLGILMANVFIYLLVEVSKNSQDKNGKGVIIIPKEKFKWPPSTPRAYWN 60
 QY 61 REQEKLRQYNPILSMILNTOTGEAGRLSNISHLNYCEPDLRVTSVVTGFFNNLPDRFKDEL 120

DB 61 REQEKLRQYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDEL 120
 QY 121 LYLRCRNYSLIDQDKCAKKPFLLLAISKLTPHFARRQAIRESWGQESNAGQTVVRVF 180
 DB 121 LYLRCRNYSLIDQDKCAKKPFLLLAISKLTPHFARRQAIRESWGQESNAGQTVVRVF 180
 QY 181 LLGOTPPEDNHPDLSMDLKFESEKHQDILMNNYRDTFFNLSLKEVLFPLRWVSTSCPDTEF 240
 DB 181 LLGOTPPEDNHPDLSMDLKFESEKHQDILMNNYRDTFFNLSLKEVLFPLRWVSTSCPDTEF 240
 QY 241 VFKGDDVFNTHILNLTNSLSKTKAKDLFIGDVIHINAGPHRDKKLYIPEVYISGLY 300
 DB 241 VFKGDDVFNTHILNLTNSLSKTKAKDLFIGDVIHINAGPHRDKKLYIPEVYISGLY 300
 QY 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLOKGLVPEKHGKGFRTDIEE 360
 DB 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLOKGLVPEKHGKGFRTDIEE 360
 QY 361 KNNKNICSYVDLMLVHSRKKPOEMIDINWSQLOSAHLKC 397
 DB 361 KNNKNICSYVDLMLVHSRKKPOEMIDINWSQLOSAHLKC 397

RESULT 10
 AAY79953
 ID AAY79953 standard; Protein: 397 AA.
 AC AAY79953;
 XX
 DT 12-MAY-2000 (first entry)
 XX
 DE Murine brainiac protein.
 XX
 KW Brainiac; egghead; cell adhesion; cytostatic; dermatological;
 KW neuroactive; cell viability; cell growth regulation; psoriasis; cancer;
 KW cell fate specification; skin lesion; nervous system defect.
 XX
 OS Mus sp.
 XX
 PN CA225126-A1.
 XX
 PD 17-JUN-1999.
 XX
 PF 17-DEC-1997; 97CA-2225126.
 PR 17-DEC-1997; 97CA-2225126.
 PA (HSCR-) HSC RES & DEV LP.
 PI Egag SE;
 XX
 DR WPI: 2000-161481/15.
 DR N-PSDB; AA288486.
 XX
 PT Mammalian EGGHEAD and BRAINIAC proteins which mediate cell to cell
 PT adhesion and may be used to treat cancer, psoriasis and other skin
 PT lesions and nervous system defects or diseases -
 XX
 PS Example 1; Page 25; 30pp; English.
 XX
 CC The present invention describes mammalian EGGHEAD and BRAINIAC proteins,
 CC which mediate cell-to-cell adhesion. The mammalian proteins are similar
 CC to the EGGHEAD and BRAINIAC genes of Drosophila melanogaster. BRAINIAC
 CC and EGGHEAD genes regulate adhesion between epithelial cells and require
 CC the presence of the Notch protein for this activity. Mutational
 CC disruption of BRAINIAC, EGGHEAD or Notch results in the loss of follicle
 CC epithelial cell adhesion, thereby reducing the efficiency of signalling
 CC through other epithelial cell receptors such as the epidermal growth
 CC factor receptor. The BRAINIAC and EGGHEAD proteins have cytostatic,
 CC dermatological and neuroactive properties. As the BRAINIAC and EGGHEAD
 CC proteins are involved in adhesion between epithelial cells, and as this
 CC adhesion is required for cell viability, cell growth regulation and cell

CC fate specification, it is envisioned that wild type or mutant forms of
CC mammalian BRAINAC and/or EGHEAD can be used to alter epithelial cell
CC adhesion. This should be useful in treating many diseases which present
CC problems of cell viability, cell growth regulation and cell fate
CC specification. For example, these proteins, or active fragments or
CC analogues of these proteins and these genes can be used to treat
CC diseases such as cancer, psoriasis and other skin lesions, and nervous
CC system defects or diseases. The present sequence represents the murine
CC BRAINAC protein.
XX
XX
SQ Sequence 397 AA;

Query Match 88.2%; Score 1872; DB 21; Length 397;
Best Local Similarity 87.4%; Pred. No. 1.2e-191;
Matches 347; Conservative 19; Mismatches 31; Indels 0; Gaps 0;
Qy 1 MSVGRRRKLLGLIMANVFYIFIMEVSKSSQKNGKEVLPKEKFWKISTPPPEAYWN 60
Db 1 MSVGRRRVKKLLGLIMANVFYILIVEVSKNSQDKNGKGVIIIPKEKFWKISTPPRAYWN 60
Qy 61 REQEKLRQYNPILSMLTNOTGEAGRLSNISHLNCEPDLRTSVVTFNNLPDRFKDFL 120
Db 61 REQEKLRWYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTFNNLPDRFKDFL 120
Qy 121 LYLRCRNYSLIIDQPKCAKKPFLLLAIKSLTPHFARROAIRSWGOESNAGNQTWVRVF 180
Db 121 LYLRCRNYSLIIDQPKCAKKPFLLLAIKSLTPHFARROAIRSWGRETNVGNQTVVRVF 180
Qy 181 LLGQTPPEDNHPDLSMDLKFESKHQDILMWNRYDTFFNLSKEVLFRLWVSTSCPDTEF 240
Db 181 LLGKTTPEDNHPDLSMDLKFESKHQDILMWNRYDTFFNLSKEVLFRLWVSTSCPDADF 240
Qy 241 VFEGDDVFNVTTHILNLYNSLSKTRAKDLFTGDIHNAHPHDKKLYIIEVYVSGLY 300
Db 241 VFEGDDVFNVTTHILNLYNSLSKTRAKDLFTGDIHNAHPHDKKLYIIEVYVSGY 300
Qy 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
Db 301 PPYAGGGFLYSGPLALRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
Qy 361 KNNKNICSYDMLVHRSRKPQEMIDIWSQLQSPNLKC 397
Db 361 KNNKNICSYDMLVHRSRKPQEMIDIWSQLQSPNLKC 397

RESULT 11
AAV69697
ID AAV69697 standard; Protein; 397 AA.
AC AAV69697;
XX
XX
DT 08-MAY-2000 (first entry)
XX
XX
DE Murine Brainiac protein.
XX
XX
KW Brainiac; murine; mammalian; expressed sequence tag; EST;
KW Drosophila melanogaster; Egghead; Notch; epithelial cell adhesion;
KW Fringe family; cell viability; growth regulation; cell fate; cancer;
KW psoriasis; skin lesion; nervous system disorder; developmental syndrome;
KW transgenic animal; drug screening.
XX
XX
OS Mus sp.
XX
XX
PN CA2255109-A1.
XX
XX
PD 17-JUN-1999.
XX
XX
PF 17-DEC-1998; 98CA-2255109.
XX
XX
PR 17-DEC-1997; 97CA-2225126.
XX
XX
PA (HSCR-) HSC RES & DEV LP.
XX

PI Egan SE;
XX
DR WPI; 2000-148082/14.
DR N-PSDB; AAZ87185.
XX
XX New nucleic acids encoding a murine and human Brainiac protein, useful
PT for detecting somatic or germline DNA lesions which are responsible for
PT developmental syndromes or diseases including cancer.
XX
XX
PS Claim 13; Page 28; 40pp; English.
XX
CC This sequence represents murine Brainiac protein. The cDNA encoding this
CC protein was isolated from a mouse mammary gland cDNA library via the use
CC of a probe generated via PCR from a variety of murine tissues. The PCR
CC primers used for probe generation (AAZ87187-287188) were based on EST
CC (expressed sequence tag) sequences with significant similarity to
CC Drosophila melanogaster Brainiac. Drosophila Brainiac and Egghead
CC proteins regulate adhesion between epithelial cells, this activity being
CC dependent on the presence of Notch. Drosophila Brainiac is a secreted
CC protein, and has sequence similarities with the Drosophila Fringe
CC proteins. Brainiac/egghead-mediated epithelial cell adhesion is
CC required for cell viability, cell growth regulation and cell fate
CC specification. Wild-type or mutant forms of mammalian Brainiac
CC proteins may therefore be used to alter epithelial cell adhesion
CC in a mammal. Mammalian Brainiac proteins, active fragments, analogues,
CC and nucleic acids may be used to treat diseases such as cancer,
CC psoriasis and other skin lesions, and nervous system disorders.
CC Mammalian Brainiac nucleic acids may also be used to detect somatic
CC or germline DNA lesions which are responsible for developmental
CC syndromes or diseases including cancer. The mammalian Brainiac proteins
CC and fragments or its analogues are useful as antigens in immunoassays
CC including enzyme-linked immunosorbent assays (ELISA), radioimmunoassays
CC (RIA) and other non-enzyme linked antibody binding assays. Non-human
CC transgenic animals comprising nucleotide sequences encoding human
CC Brainiac protein (AAV69698) can be used as animal models for the study
CC of mammalian Brainiac gene function, for the screening of candidate
CC compounds and for the evaluation of potential therapeutic interventions.
XX
SQ Sequence 397 AA;

Query Match 88.2%; Score 1872; DB 21; Length 397;
Best Local Similarity 87.4%; Pred. No. 1.2e-191;
Matches 347; Conservative 19; Mismatches 31; Indels 0; Gaps 0;
Qy 1 MSVGRRRKLLGLIMANVFYIFIMEVSKSSQKNGKEVLPKEKFWKISTPPPEAYWN 60
Db 1 MSVGRRRVKKLLGLIMANVFYILIVEVSKNSQDKNGKGVIIIPKEKFWKISTPPRAYWN 60
Qy 61 REQEKLRQYNPILSMLTNOTGEAGRLSNISHLNCEPDLRTSVVTFNNLPDRFKDFL 120
Db 61 REQEKLRWYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTFNNLPDRFKDFL 120
Qy 121 LYLRCRNYSLIIDQPKCAKKPFLLLAIKSLTPHFARROAIRSWGOESNAGNQTWVRVF 180
Db 121 LYLRCRNYSLIIDQPKCAKKPFLLLAIKSLTPHFARROAIRSWGRETNVGNQTVVRVF 180
Qy 181 LLGQTPPEDNHPDLSMDLKFESKHQDILMWNRYDTFFNLSKEVLFRLWVSTSCPDTEF 240
Db 181 LLGKTTPEDNHPDLSMDLKFESKHQDILMWNRYDTFFNLSKEVLFRLWVSTSCPDADF 240
Qy 241 VFEGDDVFNVTTHILNLYNSLSKTRAKDLFTGDIHNAHPHDKKLYIIEVYVSGLY 300
Db 241 VFEGDDVFNVTTHILNLYNSLSKTRAKDLFTGDIHNAHPHDKKLYIIEVYVSGY 300
Qy 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
Db 301 PPYAGGGFLYSGPLALRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
Qy 361 KNNKNICSYDMLVHRSRKPQEMIDIWSQLQSPNLKC 397
Db 361 KNNKNICSYDMLVHRSRKPQEMIDIWSQLQSPNLKC 397

```

RESULT 12
AAG66115
ID AAG66115 standard; Protein; 397 AA.
AC AAG66115;
XX
XX 13-MAR-2002 (first entry)
DE Murine LIG46 polypeptide.
XX
XX Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Stral3;
KW anorectic; anabolic; antisense therapy; mouse.
XX
OS Mus sp.
XX
XX Key Location/Qualifiers
XX Peptide 1..30
XX Protein /note= "signal peptide"
XX /note= "mature protein"
XX
XX US2001024808-A1.
XX
XX 27-SEP-2001.
XX
XX 12-MAR-2001; 2001US-0804357.
XX
XX 29-OCT-1998; 98US-106378P.
XX
XX 19-NOV-1998; 98US-0195896.
XX
XX 10-SEP-1998; 98US-0150857.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX White D, Zhou J, Tartaglia LA;
XX WPI: 2001-624963/72.
XX N-PSDB; AAI67865, AAI67866.
XX
XX An isolated polypeptide useful for treating a weight disorder and for
XX screening compounds that may inhibit protein activity comprises a
XX leptin-induced protein.
XX
XX Claim 22; Fig 1; 46pp; English.
XX
XX The invention relates to genes whose expression are induced by leptin.
XX LIG46 and LIG56 are novel leptin induced genes (LIG), while four other
XX genes such as Tgtp, LRG-47, RC10-II and Stral3 have been previously
XX identified. The leptin induced proteins can be expressed by standard
XX recombinant methodology. The proteins and encoding polynucleotides may
XX be used in screening assays to identify compounds that may bind to it.
XX Administering a molecule, e.g., an antisense molecule, which reduces
XX expression of activity of protein selected from LIG46, LIG56, Tgtp,
XX LRG-47, RC10-II, and Stral3 may be used to treat a weight disorder. The
XX method may further comprise administering leptin. The present sequence
XX represents a murine LIG46 protein.
XX
XX Sequence 397 AA:
Query Match 88.2%; Score 1872; DB 22; Length 397;
Best Local Similarity 87.4%; Pred. No. 1.2e-191;
Matches 347; Conservative 19; Mismatches 31; Indels 0; Gaps 0;
OY 1 MSVGRRIKLLGILMANVFIYIMEVSKSSQKNGKGVIIIPKFKFKIISTPPAYNN 60
DB 1 MSVGRRRVKLLGILMANVFIYIVEVSKNSQDKNGKGVIIIPKFKFKPPSTPRAYNN 60
OY 61 REQEKLRQXNPILSMITNCTGAGRLSNISHLNYCEPDLRTVTSVVTGNNLPDRFKDEL 120
DB 61 REQEKLRNWNPNILNRVANOTGELATSPNTSHLSYCEPDSVYMTATVDFNLPDRFKDEL 120
OY 121 LYLCRNYSLIDQPKCAKKPELLLAIKSLTPHFARROAIRSWGQESNAGNQTVVYRVF 180
DB 121 LYLCRNYSLIDQPKCAKKPELLLAIKSLIPHFARROAIRSWGRETNGVGNQTVVYRVF 180

181 ILGQTPEDNHPDLSOMLKFESKSHODILMWNYRDTFFNLSLKEVLFRLWVSTSCPDTEF 240
181 LLCKTPEDNHPDLSOMLKFESDKHODILMWNYRDTFFNLSLKEVLFRLWVSTSCPDADF 240
241 VFKGDDVFNTHHILNLSLTKAKOLFICGVTHNAGPHRDKKLKYIPEVYVYSGLY 300
241 VFKGDDVFNTHHILNLSLTKAKOLFICGVTHNAGPHRDKKLKYIPEVYVYSGLY 300
301 PPVAGGGFLYSGHLALRLVHITDQVHLVPIDOVYTGMCLOKLGVLPEKHKGFRFTDIEE 360
301 PPVAGGGFLYSGHLALRLVHITDQVHLVPIDOVYTGMCLOKLGVLPEKHKGFRFTDIEE 360
361 KNKNICSIVYDMLVHRSRKPQEMIDIWSQLOSPNLC 397
361 KNKNICSIVYDMLVHRSRKPQEMIDIWSQLOSPNLC 397

RESULT 13
AAE29092
ID AAE29092 standard; Protein; 397 AA.
XX
XX AAE29092;
XX
XX 27-JAN-2003 (first entry)
DE Murine LIG46 protein.
XX
XX LIG46; leptin; low body weight; chromosomal mapping; tissue typing;
KW forensic biology; transgenic; gene therapy; antianorectic; murine.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
XX Peptide 1..32
XX Protein /label= Signal_peptide
XX /note= "Murine mature LIG46 protein"
XX Modified-site 3..6
XX /note= "Amidation site"
XX Modified-site 30..33
XX /note= "N-glycosylation site"
XX Modified-site 31..34
XX /note= "Casein kinase II phosphorylation site"
XX Domain 33..302
XX /note= "Extracellular domain"
XX Modified-site 54..56
XX /note= "Protein kinase C phosphorylation site"
XX Modified-site 79..82
XX /note= "N-glycosylation site"
XX Modified-site 89..92
XX /note= "N-glycosylation site"
XX Modified-site 94..97
XX /note= "Casein kinase II phosphorylation site"
XX Modified-site 115..122
XX /note= "Tyrosine kinase phosphorylation site"
XX Modified-site 127..173
XX /note= "N-glycosylation site"
XX Modified-site 185..188
XX /note= "Casein kinase II phosphorylation site"
XX Modified-site 202..204
XX /note= "Protein kinase C phosphorylation site"
XX Modified-site 219..222
XX /note= "N-glycosylation site"
XX Modified-site 221..224
XX /note= "Casein kinase II phosphorylation site"
XX Modified-site 221..223
XX /note= "Protein kinase C phosphorylation site"
XX Modified-site 234..237
XX /note= "Casein kinase II phosphorylation site"
XX Domain 303..320
XX /note= "Transmembrane domain"
XX Domain 321..397

```

```
FT /note- "Cytoplasmic domain"
FT Modified-site 323..325
FT /note- "Protein kinase C phosphorylation site"
FT Modified-site 368..371
FT /note- "Casein kinase II phosphorylation site"
FT Modified-site 377..379
FT /note- "Protein kinase C phosphorylation site"
FX
FX WO200274905-A2.
FX
FX 26-SEP-2002.
FX
FX 20-NOV-2001; 2001WO-US43345.
FX
FX 21-NOV-2000; 2000US-0717778.
FX
FX (MILL-) MILLENNIUM PHARM INC.
FX
FX White DW, Zhou J, Tartaglia LA, Stricker-Krongrad A, Clausen H;
FX WPI: 2002-759885/82.
FX N-PSDB; AAD46672.
FX
FX New isolated nucleic acid molecules, designated as leptin-induced genes
FX 46 (LIG46), useful for treating a subject having a disorder
FX characterized by undesirable level of LIG46 expression or activity,
FX such as low body weight
FX
FX Claim 22; Fig 1; 90pp; English.
FX
FX The invention relates to LIG46, a gene whose expression is induced by
FX leptin. LIG46 DNA and protein are useful in treating a subject having
FX a disorder characterized by undesirable level of LIG46 expression or
FX activity, such as low body weight. They are also useful in a screening
FX assay, chromosomal mapping, tissue typing and forensic biology. The
FX probes based on the LIG46 nucleotide sequence are useful for detecting
FX transcripts or genomic sequences encoding the same or related proteins.
FX The LIG46 polypeptides are useful as immunogens for raising anti-LIG46
FX antibodies. The host cells are useful for producing non-human transgenic
FX animals. LIG46 DNA is used in gene therapy. The present sequence is
FX murine LIG46 protein.
FX
FX Sequence 397 AA:
FX
FX Query Match 88.2%; Score 1872; DB 23; Length 397;
FX Best Local Similarity 87.4%; Pred. No. 1.2e-191;
FX Matches 347; Conservative 19; Mismatches 31; Indels 0; Gaps 0;
QY 1 MSVGRRIKLLGILLMANVFYIFIMEVSKSSQKNGKGVIIPEKFKWISTPPEAYWN 60
DB 1 MSVGRRRVKLLGILLMANVFYILIVEVSKNSQDKNGKGVIIPEKFKWPPTPRAYWN 60
QY 61 REQEKLNQYNPILSLMTNQTGEAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDFL 120
DB 61 REQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSVTAVTDFNNLPDRFKDFL 120
QY 121 LYLCRCNYSLLIDOPKCAKPFLLLAIKSLTPHFARRQAIRESWGQESNAGQTVRVVF 180
DB 121 LYLCRCNYSLLIDOPKCAKPFLLLAIKSLTPHFARRQAIRESWGRTNVGNQTVRVVF 180
QY 181 LLGOTPTPDNHPDLSMDLKFESKHQDILMWNVTRDFFNLSLKEVLFRLAWVSTSCPDTEF 240
DB 181 LLGRTPTPDNHPDLSMDLKFESKHQDILMWNVTRDFFNLSLKEVLFRLWVSTSCPDAEF 240
QY 241 VFKGDDVFNTHILNLSLTKAKDLFIGDVIHNAAGPHRDKKLYIPVWVSGLY 300
DB 241 VFKGDDVFNTHILNLSLTKAKDLFIGDVIHNAAGPHRDKKLYIPVFTIGVY 300
QY 301 PPYAGGGGFLYSGHLALRLYHITDQVHLPTDDVYTCMCQKLGKLVPEKHKGFRTDIEE 360
DB 301 PPYAGGGGFLYSGPLALRLYSATSRRVHLYPDIDVYTCMCQKLGKLVPEKHKGFRTDIEE 360
QY 361 KNKNKNCISYDMLVHRSRKFQEMIDTWSQLQSPNLKC 397
```

```
DB 361 KNKNKNCISYDMLVHRSRKFQEMIDTWSQLQSPNLKC 397
DB
DB RESULT 14
DB ABB82648
DB ID ABB82648 standard; Protein; 397 AA.
DB XX ABB82648;
DB AC
DB DT 19-FEB-2003 (first entry)
DB XX
DB DE Murine beta3GnT polypeptide sequence.
DB XX
DB KW Beta3GnT; transgenic; beta-1,3-N-acetylglucosaminyltransferase;
DB disease system; mouse; enzyme.
DB OS Mus sp.
DB XX
DB PN WO200279413-A2.
DB XX
DB PD 10-OCT-2002.
DB XX
DB PF 29-MAR-2002; 2002WO-US09645.
DB XX
DB PR 29-MAR-2001; 2001US-280706P.
DB PR 28-MAR-2002; 2002US-0280706.
DB XX
DB PA (DELT-) DELTAGEN INC.
DB XX
DB PI Leviten MW, Phillips R;
DB XX
DB DR WPI: 2003-067437/06.
DB N-PSDB; ABV75081.
DB XX
DB PT New transgenic mouse comprising a disruption in a beta3GnT gene, as in
DB vivo model to study various disease states or conditions in which
DB beta3GnT may be implicated or involved, such as abnormal cell growth,
DB cancer and metastasis
DB XX
DB PS Disclosure; Fig 2; 55pp; English.
DB XX
DB CC The invention relates to a transgenic mouse comprising a disruption in a
DB beta3GnT (beta-1,3-N-acetylglucosaminyltransferase) gene, where there is
DB no native expression of beta3GnT gene. The transgenic mice may be used
DB as in vivo model to study various disease states or conditions in which
DB beta3GnT may be implicated or involved, such as abnormal cell growth,
DB cancer and metastasis, and to evaluate various treatments or to identify
DB agents for treating disease states or conditions, such as anxiety or
DB depression. Animal-based disease systems may be used to identify
DB compounds capable of ameliorating disease symptoms, as test substrates
DB for the identification of drugs, pharmaceuticals, therapies and
DB interventions that may be effective in treating a disease or other
DB phenotypic characteristic of the animal. The present sequence represents
DB a mouse beta3GnT polypeptide sequence.
DB
DB Query Sequence 397 AA:
DB
DB Query Match 87.9%; Score 1866; DB 24; Length 397;
DB Best Local Similarity 87.2%; Pred. No. 5.2e-191;
DB Matches 346; Conservative 19; Mismatches 32; Indels 0; Gaps 0;
QY 1 MSVGRRIKLLGILLMANVFYIFIMEVSKSSQKNGKGVIIPEKFKWISTPPEAYWN 60
DB 1 MSVGRRRVKLLGILLMANVFYILIVEVSKNSQDKNGKGVIIPEKFKWPPTPRAYWN 60
QY 61 REQEKLNQYNPILSLMTNQTGEAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDFL 120
DB 61 REQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSVTAVTDFNNLPDRFKDFL 120
QY 121 LYLCRCNYSLLIDOPKCAKPFLLLAIKSLTPHFARRQAIRESWGQESNAGQTVRVVF 180
DB 121 LYLCRCNYSLLIDOPKCAKPFLLLAIKSLTPHFARRQAIRESWGRTNVGNQTVRVVF 180
```


QY 181 LLGQTPEDNHPDLSMLKFESEKHQDILMWNRYDTFFNLSEKVEFLRWVSTSCPDTEF 240
DB 181 LLGKTPEDNHPDLSMLKFESEKHQDILMWNRYDTFFNLSEKVEFLRWVSTSCPDTEF 240
QY 241 VFKGDDVFNTHILNLSKTKAKDLFGDVHNNAGPHRDKKLYIPEVYSGLY 300
DB 241 VFKGDDVFNTHILNLSKTKAKDLFGDVHNNAGPHRDKKLYIPEVYSGLY 300
QY 301 PPYAGGGFLYSGHLALRYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRIFDIE 360
DB 301 PPYAGGGFLYSGHLALRYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRIFDIE 360
QY 361 KNNKNNICSYDMLVHSRKPQEMIDWSQLOSAHLKC 397
DB 361 KNNKNNICSYDMLVHSRKPQEMIDWSQLOSAHLKC 397

RESULT 15

AA79954
ID AA79954 standard; Protein: 397 AA.

XX AC AA79954;

DT 12-MAY-2000 (first entry)

XX DE Human brainiac protein.

XX KW Brainiac; egghead; cell adhesion; cytosolic; dermatological;
KW neuroactive; cell viability; cell growth regulation; psoriasis; cancer;
KW cell fate specification; skin lesion; nervous system defect.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT FT Misc-difference 1..397
FT FT /note- "any X given in the sequence is unknown and
FT FT encoded by NNN"

XX PN CA225126-A1.

XX PD 17-JUN-1999.

XX PF 17-DEC-1997; 97CA-2225126.

XX PR 17-DEC-1997; 97CA-2225126.

XX PA (HSCR-) HSC RES & DEV LP.

XX PI Egan SE;

XX WPI: 2000-161481/15.

XX DR N-PSDB; AA288487.

XX Mammalian EGGHEAD and BRAINIAC proteins which mediate cell to cell
PT adhesion and may be used to treat cancer, psoriasis and other skin
PT lesions and nervous system defects or diseases -

PS Example 2; Page 27; 30pp; English.

XX The present invention describes mammalian EGGHEAD and BRAINIAC proteins,
CC which mediate cell-to-cell adhesion. The mammalian proteins are similar
CC to the EGGHEAD and BRAINIAC genes of *Drosophila melanogaster*. BRAINIAC
CC and EGGHEAD genes regulate adhesion between epithelial cells and require
CC the presence of the Notch protein for this activity. Mutational
CC disruption of BRAINIAC, EGGHEAD or Notch results in the loss of follicle
CC epithelial cell adhesion, thereby reducing the efficiency of signalling
CC through other epithelial cell receptors such as the epidermal growth
CC factor receptor. The BRAINIAC and EGGHEAD proteins have cytosolic,
CC dermatological and neuroactive properties. As the BRAINIAC and EGGHEAD
CC proteins are involved in adhesion between epithelial cells, and as this
CC adhesion is required for cell viability, cell growth regulation and cell
CC fate specification, it is envisioned that wild type or mutant forms of

CC mammalian BRAINIAC and/or EGGHEAD can be used to alter epithelial cell
CC adhesion. This should be useful in treating many diseases which present
CC problems of cell viability, cell growth regulation and cell fate
CC specification. For example, these proteins, or active fragments or
CC analogues of these proteins and these genes can be used to treat
CC diseases such as cancer, psoriasis and other skin lesions, and nervous
CC system defects or diseases. The present sequence represents the human
XX BRAINIAC protein.

XX SQ Sequence 397 AA;

Query Match 85.9%; Score 1823; DB 21; Length 397;

Best Local Similarity 89.2%; Pred. No. 2.2e-186;

Matches 346; Conservative 2; Mismatches 40; Indels 0; Gaps 0;

QY 10 LGLIMMANVFIYFIMEVSKSSQKNGKGEVIIPKEKFKISTPPPEAYWNRQEKLNQ 69

DB 10 LGLIMMANVFIYFIMEVSKSSQKNGKGNVIIPKEKFKISTPPPYAXNRQEKLNQ 69

QY 70 YNPILSMLTNOTGEAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDELLYLRCNYS 129

DB 70 YNPILNRLTNOTGEXXXXXXXXVTFNNLPDRFKDELLYLRCNYS 129

QY 130 LLIDQPKCAKPPFLLAIAKSLTPHFARROAIRSWGQESNAGNQTVVRVFLGQTTPED 189

DB 130 LLIDQPKCAKPPFLLAIAKSLTPHFARROAIRSWGQESNAGNQTVVRVFLGQTTPED 189

QY 190 NHPDLSMLKFESEKHQDILMWNRYDTFFNLSEKVEFLRWVSTSCPDTEFVFKGDDVDF 249

DB 190 NHPDLSMLKFESEKHQDILMWNRYDTFFNLSEKVEFLRWVSTSCPDTEFVFKGDDVDF 249

QY 250 VNTHHILNLSKTKAKDLFGDVHNNAGPHRDKKLYIPEVYSGLYPYAGGGGF 309

DB 250 VNTHHILNLSKTKAKDLFGDVHNNAGPHRDKKLYIPEVYSGLYPYAGGGGF 309

QY 310 LYSGLHALRLYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRIFDIEKNNKNNICSY 369

DB 310 LYSGLHALRLYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRIFDIEKNNKNNICSY 369

QY 370 VDLMLVHSRKPQEMIDWSQLOSAHLKC 397

DB 370 VDLMLVHSRKPQEMIDWSQLOSAHLKC 397

Search completed: October 20, 2003, 23:01:59

Job time : 39.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: October 20, 2003, 22:49:49 ; Search: time 38.5 Seconds
(without alignments)
1636.739 Million cell updates/sec

Title: US-09-804-357B-2

Perfect score: 2129

Sequence: 1 MSVRRRVKLLGILMANVF.....RKPQEMDIWSQLQSPNLKC 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

```
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2119	99.5	397	21	AA1984641
2	2119	99.5	397	21	AA1984641
3	2119	99.5	397	21	AA1984641
4	2119	99.5	397	22	AA1984641
5	2119	99.5	397	23	AA1984641
6	2113	99.2	397	24	AA1984641
7	1977	92.9	367	22	AA1984641
8	1862	87.5	397	19	AA1984641
9	1862	87.5	397	21	AA1984641

10	1862	87.5	397	22	AA1984641	Human LIG46 polype
11	1862	87.5	397	23	AA1984641	Human LIG46 protei
12	1862	87.5	397	24	AA1984641	Human beta3Gnt po
13	1844	86.6	413	22	AA1984641	Human beta 1,3-N-a
14	1784	83.8	397	21	AA1984641	Human Brainiac pro
15	1753	82.3	373	22	AA1984641	Human protein sequ
16	1680	78.9	397	21	AA1984641	Human brainiac pro
17	980	46.0	197	23	AA1984641	Human Brainiac-3.
18	656	30.8	352	20	AA1984641	Human secreted pro
19	656	30.8	352	24	AA1984641	Human secreted pro
20	656	30.8	353	20	AA1984641	Human TNFR-10 pro
21	654.5	30.7	401	23	AA1984641	Human PRO4344 prot
22	647.5	30.4	378	21	AA1984641	Human beta 1,3-N-a
23	647.5	30.4	378	22	AA1984641	Human Pro peptide
24	647.5	30.4	378	23	AA1984641	The beta-1,3-galac
25	646	30.3	384	23	AA1984641	Novel human enzyme
26	635.5	29.8	378	21	AA1984641	Human galactosyltr
27	635.5	29.8	378	23	AA1984641	Human zissp6 prote
28	635.5	29.8	378	23	AA1984641	Novel human diagno
29	633.5	29.8	334	22	AA1984641	Human NOV106a prot
30	631.5	29.7	364	23	AA1984641	Human galactosyltr
31	623	29.3	415	23	AA1984641	Human beta 1,3-N-a
32	622.5	29.2	372	22	AA1984641	Human epidermoid c
33	619.5	29.1	372	19	AA1984641	Human PRO polypt
34	619.5	29.1	372	22	AA1984641	Human beta 1,3-N-a
35	619.5	29.1	372	22	AA1984641	Amino acid sequenc
36	619.5	29.1	372	23	AA1984641	Human beta1,3-N-ac
37	619.5	29.1	372	23	AA1984641	Human PRO1266 prot
38	619.5	29.1	372	24	AA1984641	Human secreted/tra
39	619.5	29.1	372	24	AA1984641	Novel human secret
40	619.5	29.1	372	24	AA1984641	Human secreted/tra
41	619.5	29.1	372	24	AA1984641	Human cancer-relat
42	619.5	29.1	372	24	AA1984641	Human PRO polypt
43	619.5	29.1	372	24	AA1984641	Human PRO polypt
44	619.5	29.1	372	24	AA1984641	Human secreted/tra
45	619.5	29.1	372	24	AA1984641	Human secreted/tra

ALIGNMENTS

RESULT 1

AA1984641
ID AA1984641 standard; Protein; 397 AA.

AC AA1984641;

DT 25-JUL-2000 (first entry)

DE Amino acid sequence of a murine LIG46 polypeptide.

KW Leptin; LIG46; body weight; leptin inducible gene; obesity; cachexia;

KW LIG56; Tgtp; LRG-47; RC10-II; Stra13.

OS Mus sp.

PN WO200015826-A2.

PO 23-MAR-2000.

PF 10-SEP-1999; 99WO-US20722.

PR 10-SEP-1998; 98US-0150857.

PR 29-OCT-1998; 98US-0106378.

PR 19-NOV-1998; 98US-0195896.

PR 15-APR-1999; 99US-0292228.

(MILL-) MILLENNIUM PHARM INC.

PI White D. Zhou J, Tartaglia LA;

XX WPI: 2000-371461/23.

DR N-PSDB; AAA12714.

```

XX  Method for determining compounds which modulate body weight and can be
PT  used to treat e.g. obesity comprises measuring the activity of leptin
PT  inducible genes
XX
XX  Claim 23: Fig 1A-B; 123pp; English.
XX
CC  The present sequence represents a leptin induced LIG46 polypeptide. The
CC  specification describes a method for determining whether a compound can
CC  be used to modulate body weight by measuring the activity of leptin
CC  inducible genes, such as LIG46. The method can be used to specifically
CC  detect LIG46 and LIG56 nucleic acid molecules relative to other nucleic
CC  acid molecules encoding galactosyltransferases or GTP-binding proteins.
CC  They can also be used in diagnostic assays to identify the presence or
CC  absence of a genetic lesion or mutation characterized by aberrant
CC  modification or mis-regulation of the genes or aberrant post-
CC  translational modification of the proteins. LIG46 and LIG56 proteins
CC  and nucleic acid molecules can be used to treat obesity or cachexia.
CC  LIG46 and LIG56 antagonists are used to treat obesity and LIG46 and
CC  LIG56 agonists are used to treat low body weight. The leptin inducible
CC  genes i.e. LIG46, LIG56, Tgfb, LRG-47, RC10-11 and Stral3 can be used
CC  to determine whether a compound modulates body weight and can then be
CC  used to treat obesity or cachexia or low body weight.
XX
SQ  Sequence 397 AA:
      Query Match 99.5%; Score 2119; DB 21; Length 397;
      Best Local Similarity 99.5%; Pred. No. 1e-223;
      Matches 395; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY  1 MSVGRRRVKLLGILMANVFIYLIVEVSKNSODKNGKGGVITPKKFKWKPSTPRAYWN 60
DB  1 MSVGRRRVKLLGILMANVFIYLIVEVSKNSODKNGKGGVITPKKFKWKPSTPRAYWN 60
QY  61 REQEKLNRWNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRKDFL 120
DB  61 REQEKLNRWNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRKDFL 120
QY  121 LYLRCRNYSLLDOPKKCAKPPFLLAISKLIPIHFAARRQAIRESWGRETNGVQTVVRF 180
DB  121 LYLRCRNYSLLDOPKKCAKPPFLLAISKLIPIHFAARRQAIRESWGRETNGVQTVVRF 180
QY  181 LLGKTPPEDNHPDLSMDLKPFESDKHQDILMWNRYDTFFNLSKEVLFLRWVSTSCDAEF 240
DB  181 LLGKTPPEDNHPDLSMDLKPFESDKHQDILMWNRYDTFFNLSKEVLFLRWVSTSCDAEF 240
QY  241 VFKGDDVFNTHILNLYNSLSKAKDLFIGDVHINAGPHRDKKLYIPEVFTGYV 300
DB  241 VFKGDDVFNTHILNLYNSLSKAKDLFIGDVHINAGPHRDKKLYIPEVFTGYV 300
QY  301 PPYAGGGFLYSGPALRLYLSATSRVHLXPIDDVYTCMLQKLGVLPEKHKGFRFTDIEE 360
DB  301 PPYAGGGFLYSGPALRLYLSATSRVHLXPIDDVYTCMLQKLGVLPEKHKGFRFTDIEE 360
QY  361 KKKKNICSYIDLMLVHSRKPQEMDINWSQLSPNLK 397
DB  361 KKKKNICSYIDLMLVHSRKPQEMDINWSQLSPNLK 397
RESULT 2
AA779953
ID  AA779953 standard; Protein; 397 AA.
XX
AC  AA779953;
XX
DT  12-MAY-2000 (first entry)
XX
DE  Murine brainiac protein.
XX
KW  Brainiac; egghead; cell adhesion; cytostatic; dermatological;
KW  neuroactive; cell viability; cell growth regulation; psoriasis; cancer;
KW  cell fate specification; skin lesion; nervous system defect.
XX

```

```

OS  Mus sp.
XX  CA2225126-A1.
XX  17-JUN-1999.
XX
XX  17-DEC-1997; 97CA-2225126.
XX
XX  17-DEC-1997; 97CA-2225126.
XX  (HSCR-) KSC RES & DEV LP.
PI  Egan SE;
XX  WPI: 2000-161481/15.
DR  N-PSDB; AA288486.
XX
PT  Mammalian EGGHEAD and BRAINIAC proteins which mediate cell to cell
PT  adhesion and may be used to treat cancer, psoriasis and other skin
PT  lesions and nervous system defects or diseases -
XX
PS  Example 1: Page 25; 30pp; English.
XX
CC  The present invention describes mammalian EGGHEAD and BRAINIAC proteins,
CC  which mediate cell-to-cell adhesion. The mammalian proteins are similar
CC  to the EGGHEAD and BRAINIAC genes of Drosophila melanogaster. BRAINIAC
CC  and EGGHEAD genes regulate adhesion between epithelial cells and require
CC  the presence of the Notch protein for this activity. Mutational
CC  disruption of BRAINIAC, EGGHEAD or Notch results in the loss of follicle
CC  epithelial cell adhesion, thereby reducing the efficiency of signalling
CC  through other epithelial cell receptors such as the epidermal growth
CC  factor receptor. The BRAINIAC and EGGHEAD proteins have cytostatic,
CC  dermatological and neuroactive properties. As the BRAINIAC and EGGHEAD
CC  proteins are involved in adhesion between epithelial cells, and as this
CC  adhesion is required for cell viability, cell growth regulation and cell
CC  fate specification, it is envisioned that wild type or mutant forms of
CC  mammalian BRAINIAC and/or EGGHEAD can be used to alter epithelial cell
CC  adhesion. This should be useful in treating many diseases which present
CC  problems of cell viability, cell growth regulation and cell fate
CC  specification. For example, these proteins, or active fragments or
CC  analogues of these proteins and these genes can be used to treat
CC  diseases such as cancer, psoriasis and other skin lesions, and nervous
CC  system defects or diseases. The present sequence represents the murine
CC  BRAINIAC protein.
XX
SQ  Sequence 397 AA:
      Query Match 99.5%; Score 2119; DB 21; Length 397;
      Best Local Similarity 99.5%; Pred. No. 1e-223;
      Matches 395; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY  1 MSVGRRRVKLLGILMANVFIYLIVEVSKNSODKNGKGGVITPKKFKWKPSTPRAYWN 60
DB  1 MSVGRRRVKLLGILMANVFIYLIVEVSKNSODKNGKGGVITPKKFKWKPSTPRAYWN 60
QY  61 REQEKLNRWNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRKDFL 120
DB  61 REQEKLNRWNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRKDFL 120
QY  121 LYLRCRNYSLLDOPKKCAKPPFLLAISKLIPIHFAARRQAIRESWGRETNGVQTVVRF 180
DB  121 LYLRCRNYSLLDOPKKCAKPPFLLAISKLIPIHFAARRQAIRESWGRETNGVQTVVRF 180
QY  181 LLGKTPPEDNHPDLSMDLKPFESDKHQDILMWNRYDTFFNLSKEVLFLRWVSTSCDAEF 240
DB  181 LLGKTPPEDNHPDLSMDLKPFESDKHQDILMWNRYDTFFNLSKEVLFLRWVSTSCDAEF 240
QY  241 VFKGDDVFNTHILNLYNSLSKAKDLFIGDVHINAGPHRDKKLYIPEVFTGYV 300
DB  241 VFKGDDVFNTHILNLYNSLSKAKDLFIGDVHINAGPHRDKKLYIPEVFTGYV 300
QY  301 PPYAGGGFLYSGPALRLYLSATSRVHLXPIDDVYTCMLQKLGVLPEKHKGFRFTDIEE 360
DB  301 PPYAGGGFLYSGPALRLYLSATSRVHLXPIDDVYTCMLQKLGVLPEKHKGFRFTDIEE 360

```

Db 301 PPYAGGGFLYSGPLALRLYSATSRVHLYPIDDVYTGMCLOKGLGLVPEKHKGFRFTFIDE 360

Qy 361 KNKKNICSYIDLMLVHSRKPQEMIDIWSQLSPNLKC 397
 |||||
 Db 361 KNKKNICSYIDLMLVHSRKPQEMIDIWSQLSPNLKC 397

RESULT 3
 ID AAY69697
 AC AAY69697 standard; Protein: 397 AA.
 XX AAY69697;
 XX 08-MAY-2000 (first entry)
 XX Murine Brainiac protein.
 XX Brainiac; murine; mammalian; expressed sequence tag; EST;
 KW Drosophila melanogaster; Egghead; Notch; epithelial cell adhesion;
 KW Fringe family; cell viability; growth regulation; cell fate; cancer;
 KW Psoriasis; skin lesion; nervous system disorder; developmental syndrome;
 KW transgenic animal; drug screening.
 XX Mus sp.
 XX CA2255109-A1.
 XX 17-JUN-1999.
 XX 17-DEC-1998; 98CA-2255109.
 XX 17-DEC-1997; 97CA-2225126.
 XX (HSCR-) HSC RES & DEV LP.
 PA Egan SE;
 PI WPI; 2000-148082/14.
 XX N-PSDB; AAZ87185.
 XX New nucleic acids encoding a murine and human Brainiac protein, useful
 PT for detecting somatic or germline DNA lesions which are responsible for
 PT developmental syndromes or diseases including cancer -
 XX Claim 13; Page 28; 40pp; English.
 XX This sequence represents murine Brainiac protein. The cDNA encoding this
 CC protein was isolated from a mouse mammary gland cDNA library via the use
 CC of a probe generated via PCR from a variety of murine tissues. The PCR
 CC primers used for probe generation (AAZ87187-287188) were based on EST
 CC (expressed sequence tag) sequences with significant similarity to
 CC Drosophila melanogaster Brainiac. Drosophila Brainiac and Egghead
 CC proteins regulate adhesion between epithelial cells, this activity being
 CC dependent on the presence of Notch. Drosophila Brainiac is a secreted
 CC protein, and has sequence similarities with the Drosophila Fringe
 CC proteins. Brainiac/Egghead-mediated epithelial cell adhesion is
 CC required for cell viability, cell growth regulation and cell fate
 CC specification. Wild-type or mutant forms of mammalian Brainiac
 CC proteins may therefore be used to alter epithelial cell adhesion
 CC in a mammal. Mammalian Brainiac proteins, active fragments, analogues,
 CC and nucleic acids may be used to treat diseases such as cancer,
 CC psoriasis and other skin lesions, and nervous system disorders.
 CC Mammalian Brainiac nucleic acids may also be used to detect somatic
 CC or germline DNA lesions which are responsible for developmental
 CC syndromes or diseases including cancer. The mammalian Brainiac proteins
 CC and fragments or its analogues are useful as antigens in immunoassays
 CC including enzyme-linked immunosorbent assays (ELISA), radioimmunoassays
 CC (RIA) and other non-enzyme linked antibody binding assays. Non-human
 CC transgenic animals comprising nucleotide sequences encoding human
 CC Brainiac protein (AA69698) can be used as animal models for the study
 CC of mammalian Brainiac gene function, for the screening of candidate
 CC compounds and for the evaluation of potential therapeutic interventions.
 XX

SQ Sequence 397 AA:
 Query Match 99.5%; Score 2119; DB 21; Length 397;
 Best Local Similarity 99.5%; Pred. No. 1e-223; Indels 0; Gaps 0;
 Matches 395; Conservative 0; Mismatches 2;

Qy 1 MSVGRRRVKLLGILMMANVFYLIIVEVSKNSQDKNGKGGVLIIPKEKFWKPPSTPRAYWN 60
 |||||
 Db 1 MSVGRRRVKLLGILMMANVFYLIIVEVSKNSQDKNGKGGVLIIPKEKFWKPPSTPRAYWN 60
 |||||
 Qy 61 REQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDFL 120
 |||||
 Db 61 REQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDFL 120
 |||||
 Qy 121 LYLRGRNYSLLIDQPKCAKPFLLAIAKSLIPHARRQAIRRESWGRETNYGNOTVVRVF 180
 |||||
 Db 121 LYLRGRNYSLLIDQPKCAKPFLLAIAKSLIPHARRQAIRRESWGRETNYGNOTVVRVF 180
 |||||
 Qy 181 LLGKTPPENHPDLSDMLKFESDKHODILMMNYRDTFFNLSLKEVFLRWVSTSCPDADF 240
 |||||
 Db 181 LLGKTPPENHPDLSDMLKFESDKHODILMMNYRDTFFNLSLKEVFLRWVSTSCPDADF 240
 |||||
 Qy 241 VFKGDDVFNTHILNYLSLSKSKAKDLFTGDVIHNAGPHRDKKLYIPEVFTGY 300
 |||||
 Db 241 VFKGDDVFNTHILNYLSLSKSKAKDLFTGDVIHNAGPHRDKKLYIPEVFTGY 300
 |||||
 Qy 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGFRFTFIDE 360
 |||||
 Db 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGFRFTFIDE 360
 |||||
 Qy 361 KNKKNICSYIDLMLVHSRKPQEMIDIWSQLSPNLKC 397
 |||||
 Db 361 KNKKNICSYIDLMLVHSRKPQEMIDIWSQLSPNLKC 397
 |||||

RESULT 4
 AAG66115
 ID AAG66115 standard; Protein: 397 AA.
 XX AAG66115;
 XX 13-MAR-2002 (first entry)
 XX Murine LIG46 polypeptide.
 XX Leptin; LIG46; leptin induced gene; Tgtp: LRG-47; RC10-II; Stral3;
 KW anorectic; anabolic; antisense therapy; mouse.
 XX Mus sp.
 FH Key Location/Qualifiers
 FT Peptide 1..30
 FT Protein /note= "signal peptide"
 FT Protein 31..397
 FT Protein /note= "mature protein"
 XX US2001024808-A1.
 XX 27-SEP-2001.
 XX 12-MAR-2001; 2001US-0804357.
 XX 29-OCT-1998; 98US-106378P.
 XX 19-NOV-1998; 98US-0195896.
 XX 10-SEP-1998; 98US-0150857.
 XX (MILL-) MILLENNIUM PHARM INC.
 XX White D, Zhou J, Tartaglia LA;
 XX WPI; 2001-624963/72.
 DR N-PSDB; AAI67865, AAI67866.
 XX

PT An isolated polypeptide useful for treating a weight disorder and for
 PT screening compounds that may inhibit protein activity comprises a
 PT leptin-induced protein

PS Claim 22: Fig 1: 46pp; English.

CC The invention relates to genes whose expression are induced by leptin.
 CC LIG46 and LIG56 are novel leptin induced genes (LIG), while four other
 CC genes such as Tgtp, LRG-47, RC10-II and Stral3 have been previously
 CC identified. The leptin induced proteins can be expressed by standard
 CC recombinant methodology. The proteins and encoding polynucleotides may
 CC be used in screening assays to identify compounds that may bind to it.
 CC Administering a molecule, e.g., an antisense molecule, which reduces
 CC expression of activity of protein selected from LIG46, LIG56, Tgtp,
 CC LRP-47, RC10-II, and Stral3 may be used to treat a weight disorder. The
 CC method may further comprise administering leptin. The present sequence
 CC represents a murine LIG46 protein.

XX Sequence 397 AA;

Query Match 99.5%; Score 2119; DB 22: Length 397;

Best Local Similarity 99.5%; Pred. No. 1e-223;

Matches 395; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSVGRRVKLLGLIMANVFYILIVEVSKNSQDKNGGVIIPKEKFKWPPSPRAYWN 60
 DB 1 MSVGRRVKLLGLIMANVFYILIVEVSKNSQDKNGGVIIPKEKFKWPPSPRAYWN 60
 QY 61 REEKLNRWYNPILNRVANOTGELATSPNTHLSYCEPDSTVMTAVTDFNLPDRFKDFL 120
 DB 61 REEKLNRWYNPILNRVANOTGELATSPNTHLSYCEPDSTVMTAVTDFNLPDRFKDFL 120
 QY 121 LYLRCRNYSLIDQPKKCAKPFLLAIKSLIPHARRQAIRFSGRETNYGNOTVVRVF 180
 DB 121 LYLRCRNYSLIDQPKKCAKPFLLAIKSLIPHARRQAIRFSGRETNYGNOTVVRVF 180
 QY 181 LKATPEDNHPDLSMLKFESKHODILMNYRTFFNLSLKEYFLFLRWYSTSCPAEF 240
 DB 181 LKATPEDNHPDLSMLKFESKHODILMNYRTFFNLSLKEYFLFLRWYSTSCPAEF 240
 QY 241 VFKGDDVFNTHIILNLSLSKSKAKDLFIGDVIHNAHPHRRKLLKYYIPEVFYGVY 300
 DB 241 VFKGDDVFNTHIILNLSLSKSKAKDLFIGDVIHNAHPHRRKLLKYYIPEVFYGVY 300
 QY 301 PPVAGGGFLYSGFALLRLYSATSRVHLVPIDVVTGMCLOKGLVPEKHKGFRTDIEE 360
 DB 301 PPVAGGGFLYSGFALLRLYSATSRVHLVPIDVVTGMCLOKGLVPEKHKGFRTDIEE 360
 QY 361 KNKKNICSYIDLMVHSRKPQEMID:WSQLQSPNLKC 397
 DB 361 KNKKNICSYIDLMVHSRKPQEMID:WSQLQSPNLKC 397

RESULT 5

AAE29092

ID AAE29092 standard; Protein: 397 AA.

XX AC AAE29092;

XX 27-JAN-2003 (first entry)

XX Murine LIG46 protein.

KW LIG46: leptin; low body weight; chromosomal mapping; tissue typing;
 KW forensic biology; transgenic; gene therapy; antianorectic; murine.

XX Mus sp.

XX Key Location/Qualifiers

XX 1..32 /label= Signal_peptide

XX 33..397 /note= "Murine mature LIG46 protein"

FT Modified-site 3..6 "Amidation site"
 FT Modified-site 30..33 "N-glycosylation site"
 FT Modified-site 31..34 "Casein kinase II phosphorylation site"
 FT Domain 33..302 "Extracellular domain"
 FT Modified-site 54..56 "Protein kinase C phosphorylation site"
 FT Modified-site 79..82 "N-glycosylation site"
 FT Modified-site 89..92 "N-glycosylation site"
 FT Modified-site 94..97 "Casein kinase II phosphorylation site"
 FT Modified-site 115..122 "Tyrosine kinase phosphorylation site"
 FT Modified-site 127..173 "N-glycosylation site"
 FT Modified-site 185..188 "Casein kinase II phosphorylation site"
 FT Modified-site 202..204 "Protein kinase C phosphorylation site"
 FT Modified-site 219..222 "N-glycosylation site"
 FT Modified-site 221..224 "Casein kinase II phosphorylation site"
 FT Modified-site 234..237 "Protein kinase C phosphorylation site"
 FT Domain 303..320 "Casein kinase II phosphorylation site"
 FT Domain 321..397 "Transmembrane domain"
 FT Modified-site 323..325 "Cytoplasmic domain"
 FT Modified-site 368..371 "Protein kinase C phosphorylation site"
 FT Modified-site 377..379 "Casein kinase II phosphorylation site"
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 PN W0200274905-A2.
 XX 26-SEP-2002.
 XX 20-NOV-2001; 2001WO-US43345.
 XX 21-NOV-2000; 2000US-0717778.
 XX (MILLI-) MILLENNIUM PHARM INC.
 XX White DW, Zhou J, Tartaglia LA, Stricker-krongrad A, Clausen H;
 DR WPI: 2002-759886/82.
 DR N-PSDB; AAD46672.
 XX New isolated nucleic acid molecules, designated as leptin-induced genes
 PT 46 (LIG46), useful for treating a subject having a disorder
 PT characterized by undesirable level of LIG46 expression or activity,
 PT such as low body weight
 XX Claim 22: Fig 1; 90pp; English.
 PS The invention relates to LIG46, a gene whose expression is induced by
 CC leptin. LIG46 DNA and protein are useful in treating a subject having
 CC a disorder characterized by undesirable level of LIG46 expression or
 CC activity, such as low body weight. They are also useful in a screening
 CC assay, chromosomal mapping, tissue typing and forensic biology. The
 CC probes based on the LIG46 nucleotide sequence are useful for detecting
 CC transcripts or genomic sequences encoding the same or related proteins.
 CC The LIG46 polypeptides are useful as immunogens for raising anti-LIG46

CC antibodies. The host cells are useful for producing non-human transgenic
 CC animals. LIG46 DNA is used in gene therapy. The present sequence is
 CC murine LIG46 protein.
 XX
 SQ Sequence 397 AA;
 Query Match 99.5%; Score 2119; DB 23; Length 397;
 Best Local Similarity 99.5%; Pred. No. 1e-223;
 Matches 395; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSVGRRRVKLLGILMANVFIYLIIVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYWN 60
 DB 1 MSVGRRRVKLLGILMANVFIYLIIVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYWN 60
 QY 61 REQEKLNRYNPILNRVANQTGELATSPNTSHLSYCEPSTVMTAVTDFNLPDRFKDFL 120
 DB 61 REQEKLNRYNPILNRVANQTGELATSPNTSHLSYCEPSTVMTAVTDFNLPDRFKDFL 120
 QY 121 LYLRCRNYSLLDQPKKCAKPPFLLLAIAKSLIPHARRQAIRESWGRETNGNOTVVRVF 180
 DB 121 LYLRCRNYSLLDQPKKCAKPPFLLLAIAKSLIPHARRQAIRESWGRETNGNOTVVRVF 180
 QY 181 LLGKTPPEDNHPDLSMDLKFESEKHODILMWNRYDTFFNLSEKVLFLRWVSTSCPDADF 240
 DB 181 LLGKTPPEDNHPDLSMDLKFESEKHODILMWNRYDTFFNLSEKVLFLRWVSTSCPDADF 240
 QY 241 VFKGDDVDFVNTTHILNLSKSKAKOLFIDGVIHNAGPHRDKKLKYYIPEVFTGYV 300
 DB 241 VFKGDDVDFVNTTHILNLSKSKAKOLFIDGVIHNAGPHRDKKLKYYIPEVFTGYV 300
 QY 301 PPVAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGRTFDIEE 360
 DB 301 PPVAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGRTFDIEE 360
 QY 361 KNKKNICSYIDLMLVHSRKPQEMIDIWSQLQSPNLKC 397
 DB 361 KNKKNICSYIDLMLVHSRKPQEMIDIWSQLQSPNLKC 397
 RESULT 6
 ABB82648
 ID ABB82648 standard; Protein: 397 AA.
 XX
 AC ABB82648;
 XX
 DT 19-FEB-2003 (first entry)
 DE Murine beta3Gnt polypeptide sequence.
 DE Beta3Gnt; transgenic; beta-1,3-N-acetylglucosaminyltransferase;
 KW disease system; mouse; enzyme.
 XX
 OS Mus sp.
 XX
 WO20027943-A2.
 XX
 PD 10-OCT-2002.
 XX
 EF 29-MAR-2002; 2002WO-US09645.
 XX
 PR 29-MAR-2001; 2001US-280706P.
 PR 28-MAR-2002; 2002US-0280706.
 XX
 XX (DELT-) DELTAGEN INC.
 PA Leventen MW, Phillips R;
 PI WPI; 2003-067437/06.
 XX N-PSDB; ABV75081.
 DR
 DR
 XX New transgenic mouse comprising a disruption in a beta3Gnt gene, as in
 PT vivo model to study various disease states or conditions in which
 PT beta3Gnt may be implicated or involved, such as abnormal cell growth,

PT cancer and metastasis -
 XX Disclosure; Fig 2; 55pp; English.
 XX
 CC The invention relates to a transgenic mouse comprising a disruption in a
 CC beta3Gnt (beta-1,3-N-acetylglucosaminyltransferase) gene, where there is
 CC no native expression of beta3Gnt gene. The transgenic mice may be used
 CC as in vivo model to study various disease states or conditions in which
 CC beta3Gnt may be implicated or involved, such as abnormal cell growth,
 CC cancer and metastasis, and to evaluate various treatments or to identify
 CC agents for treating disease states or conditions, such as anxiety or
 CC depression. Animal-based disease systems may be used to identify
 CC compounds capable of ameliorating disease symptoms, as test substrates
 CC for the identification of drugs, pharmaceuticals, therapies and
 CC interventions that may be effective in treating a disease or other
 CC phenotypic characteristic of the animal. The present sequence represents
 CC a mouse beta3Gnt polypeptide sequence.
 XX
 SQ Sequence 397 AA;
 Query Match 99.2%; Score 2113; DB 24; Length 397;
 Best Local Similarity 99.2%; Pred. No. 4.8e-223;
 Matches 394; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSVGRRRVKLLGILMANVFIYLIIVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYWN 60
 DB 1 MSVGRRRVKLLGILMANVFIYLIIVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYWN 60
 QY 61 REQEKLNRYNPILNRVANQTGELATSPNTSHLSYCEPSTVMTAVTDFNLPDRFKDFL 120
 DB 61 REQEKLNRYNPILNRVANQTGELATSPNTSHLSYCEPSTVMTAVTDFNLPDRFKDFL 120
 QY 121 LYLRCRNYSLLDQPKKCAKPPFLLLAIAKSLIPHARRQAIRESWGRETNGNOTVVRVF 180
 DB 121 LYLRCRNYSLLDQPKKCAKPPFLLLAIAKSLIPHARRQAIRESWGRETNGNOTVVRVF 180
 QY 181 LLGKTPPEDNHPDLSMDLKFESEKHODILMWNRYDTFFNLSEKVLFLRWVSTSCPDADF 240
 DB 181 LLGKTPPEDNHPDLSMDLKFESEKHODILMWNRYDTFFNLSEKVLFLRWVSTSCPDADF 240
 QY 241 VFKGDDVDFVNTTHILNLSKSKAKOLFIDGVIHNAGPHRDKKLKYYIPEVFTGYV 300
 DB 241 VFKGDDVDFVNTTHILNLSKSKAKOLFIDGVIHNAGPHRDKKLKYYIPEVFTGYV 300
 QY 301 PPVAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGRTFDIEE 360
 DB 301 PPVAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGRTFDIEE 360
 QY 361 KNKKNICSYIDLMLVHSRKPQEMIDIWSQLQSPNLKC 397
 DB 361 KNKKNICSYIDLMLVHSRKPQEMIDIWSQLQSPNLKC 397
 RESULT 7
 AAG66116
 ID AAG66116 standard; Protein: 367 AA.
 XX
 AC AAG66116;
 XX
 DT 13-MAR-2002 (first entry)
 DE Murine LIG46 mature polypeptide sequence.
 XX
 KW Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Stra13;
 KW anorectic; anabolic; antisense therapy; mouse.
 XX
 OS Mus sp.
 XX
 US2001024808-A1.
 PN
 XX 27-SEP-2001.
 PD
 XX 12-MAR-2001; 2001US-0804357.
 PF

```

XX 29-OCT-1998: 98US-106378P.
PR 19-NOV-1998: 98US-0195096.
PR 10-SEP-1998: 98US-0150857.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX White D, Zhou J, Tartaglia LA;
XX WPT: 2001-624963/72.
XX N-PSDB; AA167865, AA167866.
XX
XX An isolated polypeptide useful for treating a weight disorder and for
XX screening compounds that may inhibit protein activity comprises a
XX lepton-induced protein -
XX
XX Claim 22; Fig 1; 46pp; English.
XX
XX The invention relates to genes whose expression are induced by leptin.
XX LIG46 and LIG56 are novel leptin induced genes (LIG), while four other
XX genes such as Tgtp, LRG-47, RC10-II and Stral3 have been previously
XX identified. The leptin induced proteins can be expressed by standard
XX recombinant methodology. The proteins and encoding polynucleotides may
XX be used in screening assays to identify compounds that may bind to it.
XX Administering a molecule, e.g., an antisense molecule, which reduces
XX expression of activity of protein selected from LIG46, LIG56, Tgtp,
XX LRG-47, RC10-II, and Stral3 may be used to treat a weight disorder. The
XX method may further comprise administering leptin. The present sequence
XX represents a murine LIG46 mature protein.
XX
XX Sequence 367 AA;
XX
XX Query Match 92.9%; Score 1977; DB 22; Length 367;
XX Best Local Similarity 99.9%; Pred. No. 3.8e-208;
XX Matches 365; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
Qy 31 SSQDKNGKGGVITPEKFKKPPSTPRAYWNRQEKLNRYNPILNRVANOTGELATSPNT 90
Db 1 SSQDKNGKGGVITPEKFKKPPSTPRAYWNRQEKLNRYNPILNRVANOTGELATSPNT 60
Qy 91 SHLSYCEPDSVTMTAVTDENNLDPDKFLLYLCRNYSLLDIDQPKKAKKPFLLAIAKS 150
Db 61 SHLSYCEPDSVTMTAVTDENNLDPDKFLLYLCRNYSLLDIDQPKKAKKPFLLAIAKS 120
Qy 151 LIPHFARQAIRESWGRETNGVQTVRVVFLGKTPPEDNHPDLSMDLKFESDKHQDITLM 210
Db 121 LIPHFARQAIRESWGRETNGVQTVRVVFLGKTPPEDNHPDLSMDLKFESDKHQDITLM 180
Qy 211 WNYRDTFFNLKLEVFLRWVSTSCPDAEVFVKGGDDVFNTHILNLYNLSKSKAKDL 270
Db 181 WNYRDTFFNLKLEVFLRWVSTSCPDAEVFVKGGDDVFNTHILNLYNLSKSKAKDL 240
Qy 271 FIGDVHINAGPHRDKKLYIPEVFTGYVPPYAGGGFLYSGPALLRLYSATSRVHLYP 330
Db 241 FIGDVHINAGPHRDKKLYIPEVFTGYVPPYAGGGFLYSGPALLRLYSATSRVHLYP 300
Qy 331 IDDVYTCMCKLGLVPEKHKGFRFTDIEKKNKNCISYIDMLVHRSKRPQEMIDINSOL 390
Db 301 IDDVYTCMCKLGLVPEKHKGFRFTDIEKKNKNCISYIDMLVHRSKRPQEMIDINSOL 360
Qy 391 QSPNLKC 397
Db 361 QSPNLKC 367
XX
XX RESULT 8
XX AA80212
XX ID AA80212 standard; Protein: 397 AA.
XX
XX AC AA80212;
XX
XX DT 18-JAN-1999 (first entry)
XX

```

```

DE Human cardiac and pancreatic protein (CAPP).
XX
XX Cardiac and pancreatic protein: CAPP; muscle-derived growth factor;
XX human: pancreatitis; myocardial infarction; cardiomyopathy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..32
XX Protein 33..397
XX Mat_protein
XX Peptide 1..11
XX Peptide 29..72
XX Peptide 78..89
XX Peptide 94..105
XX Peptide 110..119
XX Peptide 124..142
XX Peptide 151..176
XX Peptide 184..218
XX Peptide 232..251
XX Peptide 262..272
XX Peptide 280..290
XX Peptide 346..368
XX Peptide 376..385
XX
XX W09844112-A1.
XX
XX 08-OCT-1998.
XX
XX 27-MAR-1998: 98WO-US06022.
XX
XX 28-MAR-1997: 97US-0042855.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Soppet DR;
XX WPI: 1998-557111/47.
XX N-PSDB; AAV66367.
XX
XX New isolated cardiac and pancreatic protein - used to develop
XX products for the diagnosis and treatment of e.g. pancreatitis or
XX abnormal hypertrophy of the heart
XX
XX Claim 1; Fig 1; 113pp; English.
XX
XX This is the amino acid sequence of novel cardiac and pancreatic
XX protein (CAPP), a member of the muscle-derived growth factor
XX superfamily, as deduced from the nucleotide sequence (see AAV66367)
XX of a cDNA clone discovered in a cDNA library derived from activated
XX T cells. The gene is abundant in adult heart and pancreas. CAPP
XX isolated nucleic acid molecules and polypeptides are provided, as
XX are vectors, host cells and recombinant methods for producing the
XX same. The invention further relates to screening methods for
XX identifying agonists and antagonists of CAPP activity, as well as
XX diagnostic methods for detecting the presence of activated T-cells,
XX and mature heart, pancreas and placental tissues and cells. The
XX CAPP polypeptides can modulate the differentiation and proliferation
XX of cells and tissue, both in vivo and ex vivo. The products can be
XX

```

CC used in the diagnosis and treatment of pancreatitis and conditions
 CC that cause abnormal hypertrophy of the heart, such as hypertension,
 CC myocardial infarction, valve disease and cardiomyopathy. The
 CC products can also be used in detection and cell culturing.

XX
 SQ Sequence 397 AA;

Query Match 87.5%; Score 1862; DB 19; Length 397;
 Best Local Similarity 86.9%; Pred. No. 1.9e-195;
 Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;

QY 1 MSVGRRRVLLGLIMMANVFIYLIIVSVKNSQDKNGKGVIIIPKEKFWKPPSIPRAYWN 60
 DB 1 MSVGRRRVLLGLIMMANVFIYFIMEVSKSSQEKNGKGVIIIPKEKFWKPPSIPRAYWN 60

QY 61 RQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDNNLPDRFKDFL 120
 DB 61 RQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDNNLPDRFKDFL 120

QY 121 LYLCRNYSLLIDOPKCAKPPFLLAIKSLIPHFARQAIRESGRETNVGNQTVVRVF 180
 DB 121 LYLCRNYSLLIDOPKCAKPPFLLAIKSLIPHFARQAIRESGRETNVGNQTVVRVF 180

QY 181 LLGKTPPEDNHPDLSMLKFESDKHODILMNNYRTDFFNLSLKEVLFRLRWSTSCPDADF 240
 DB 181 LLGKTPPEDNHPDLSMLKFESDKHODILMNNYRTDFFNLSLKEVLFRLRWSTSCPDADF 240

QY 241 VFKGDDVFNTHHILNLYNSLSKSKAKDLFGDVIHNAAGPHRDKLKYIPEVFTGVY 300
 DB 241 VFKGDDVFNTHHILNLYNSLSKSKAKDLFGDVIHNAAGPHRDKLKYIPEVFTGVY 300

QY 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
 DB 301 PPYAGGGGFLYSGHLLRLYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360

QY 361 KKKNNICSYDILMLVHSRKPQEMIDWISQLOSPNLK 397
 DB 361 KKKNNICSYDILMLVHSRKPQEMIDWISQLOSAHLK 397

RESULT 10
 AAG66118
 ID AAG66118 standard; Protein; 397 AA.
 XX
 AC AAG66118;
 XX
 DT 13-MAR-2002 (first entry)
 XX
 DE Human LIG46 polypeptide.
 XX
 KW Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Stral3;
 KW anorectic; anabolic; antisense therapy; human.
 XX
 OS Homo sapiens.
 XX

Method for determining compounds which modulate body weight and can be used to treat e.g. obesity comprises measuring the activity of leptin inducible genes -

Disclosure; Fig 8; 123pp; English.

The present sequence represents a leptin induced LIG46 polypeptide. The specification describes a method for determining whether a compound can be used to modulate body weight by measuring the activity of leptin inducible genes, such as LIG46. The method can be used to specifically detect LIG46 and LIG56 nucleic acid molecules relative to other nucleic acid molecules encoding galactosyltransferases or GTP-binding proteins. They can also be used in diagnostic assays to identify the presence or absence of a genetic lesion or mutation characterized by aberrant modification or mis-regulation of the genes or aberrant post-translational modification of the proteins. LIG46 and LIG56 proteins and nucleic acid molecules can be used to treat obesity or cachexia. LIG46 and LIG56 antagonists are used to treat obesity and LIG46 and LIG56 agonists are used to treat low body weight. The leptin inducible genes i.e. LIG46, LIG56, Tgtp, LRG-47, RC10-II and Stral3 can be used to determine whether a compound modulates body weight and can then be used to treat obesity or cachexia or low body weight.

Query Match 87.5%; Score 1862; DB 21; Length 397;
 Best Local Similarity 86.9%; Pred. No. 1.9e-195;
 Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;

QY 1 MSVGRRRVLLGLIMMANVFIYLIIVSVKNSQDKNGKGVIIIPKEKFWKPPSIPRAYWN 60
 DB 1 MSVGRRRVLLGLIMMANVFIYFIMEVSKSSQEKNGKGVIIIPKEKFWKPPSIPRAYWN 60

QY 61 RQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDNNLPDRFKDFL 120
 DB 61 RQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDNNLPDRFKDFL 120

QY 121 LYLCRNYSLLIDOPKCAKPPFLLAIKSLIPHFARQAIRESGRETNVGNQTVVRVF 180
 DB 121 LYLCRNYSLLIDOPKCAKPPFLLAIKSLIPHFARQAIRESGRETNVGNQTVVRVF 180

QY 181 LLGKTPPEDNHPDLSMLKFESDKHODILMNNYRTDFFNLSLKEVLFRLRWSTSCPDADF 240
 DB 181 LLGKTPPEDNHPDLSMLKFESDKHODILMNNYRTDFFNLSLKEVLFRLRWSTSCPDADF 240

QY 241 VFKGDDVFNTHHILNLYNSLSKSKAKDLFGDVIHNAAGPHRDKLKYIPEVFTGVY 300
 DB 241 VFKGDDVFNTHHILNLYNSLSKSKAKDLFGDVIHNAAGPHRDKLKYIPEVFTGVY 300

QY 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
 DB 301 PPYAGGGGFLYSGHLLRLYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360

QY 361 KKKNNICSYDILMLVHSRKPQEMIDWISQLOSPNLK 397
 DB 361 KKKNNICSYDILMLVHSRKPQEMIDWISQLOSAHLK 397

RESULT 9
 AAY84683
 ID AAY84683 standard; Protein; 397 AA.
 XX
 AC AAY84683;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of a human LIG46 polypeptide.
 XX
 KW Leptin; LIG46; body weight; leptin inducible gene; obesity; cachexia;
 KW LIG56; Tgtp; LRG-47; RC10-II; Stral3.
 XX
 OS Homo sapiens.
 XX
 PN WO200015826-A2.
 XX
 XX 23-MAR-2000.
 XX
 PF 10-SEP-1999; 99WO-US20722.
 XX
 PR 10-SEP-1998; 98US-0150857.
 PR 29-OCT-1998; 98US-0106378.
 PR 19-NOV-1998; 98US-0195896.
 PR 13-APR-1999; 99US-0292228.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI White D, Zhou J, Tartaglia LA;
 XX
 DR WPI; 2000-271461/23.
 DR N-PSDB; AAA12716.
 XX

FH Key Location:/Qualifiers
FT Peptide 1..32
FT Protein /note= "signal peptide"
FT Protein 33..397
FT Protein /note= "mature protein"
XX US2001024808-A1.
XX 27-SEP-2001.
XX 12-MAR-2001; 2001US-0804357.
XX 29-OCT-1998; 98US-106378P.
XX 19-NOV-1998; 98US-0195896.
XX 10-SEP-1998; 98US-0150837.
XX (MILL-) MILLENNIUM PHARM INC.
XX White D, Zhou J, Tartaglia LA;
XX WPI: 2001-624963/72.
XX N-PSDB; AA167869.
XX An isolated polypeptide useful for treating a weight disorder and for screening compounds that may inhibit protein activity comprises a leptin-induced protein.
XX Example 2; Fig 8; 46pp; English.
XX The invention relates to genes whose expression are induced by leptin. Lig46 and Lig56 are novel leptin induced genes (LIG), while four other genes such as Tgtp, LRG-47, RC10-11 and Stral3 have been previously identified. The leptin induced proteins can be expressed by standard recombinant methodology. The proteins and encoding polynucleotides may be used in screening assays to identify compounds that may bind to it. Administering a molecule, e.g., an antisense molecule, which reduces expression of activity of protein selected from LIG46, LIG56, Tgtp, LRG-47, RC10-11, and Stral3 may be used to treat a weight disorder. The method may further comprise administering leptin. The present sequence represents a human LIG46 protein.
XX Sequence 397 AA:
Query Match 87.5%; Score 1862; DB 22; Length 397;
Best Local Similarity 86.9%; Pred. No. 1.9e-195;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;
QY 1 MSVGRRRVKKLGLMMANVFYILIVEVSKNSODKNGKGVIIIPKEFKWPPSTPAYWN 60
DB 1 MSVGRRRVKKLGLMMANVFYILIVEVSKNSODKNGKGVIIIPKEFKWPPSTPAYWN 60
QY 61 REQEKLNWYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRKDFL 120
DB 61 REQEKLNWYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRKDFL 120
QY 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLIPHARQATRESWGRETWVGVVYRVF 180
DB 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLIPHARQATRESWGRETWVGVVYRVF 180
QY 181 LLGKTPEDNHPDLSMDLKFESDKHQDILMNNYRTFFNLKLVFLRWYSTSCPAEF 240
DB 181 LLGKTPEDNHPDLSMDLKFESDKHQDILMNNYRTFFNLKLVFLRWYSTSCPAEF 240
QY 241 VFKGDDVFNTHHILNLYNSLSKAKDLFTGVDVHNAGPHRDKLKYIPEYFYGVY 300
DB 241 VFKGDDVFNTHHILNLYNSLSKAKDLFTGVDVHNAGPHRDKLKYIPEYFYGVY 300
QY 301 PPVAGGGGLYSGPALLRLYTSRVHLYPIDDTVTGMCLOKGLVPEKHKGFTFDIEE 360
DB 301 PPVAGGGGLYSGHALLRLYHTDQVHLYPIDDTVTGMCLOKGLVPEKHKGFTFDIEE 360
QY 361 KNNKNTCSYDLMVHRSRKPQEMDIWSOLQSAHLKC 397

DB 361 KNNKNTCSYDLMVHRSRKPQEMDIWSOLQSAHLKC 397
RESULT 11
AAE29093
ID AAE29093 standard; Protein; 397 AA.
XX AAE29093;
XX 27-JAN-2003 (first entry)
XX Human LIG46 protein.
XX LIG46; leptin; low body weight; chromosomal mapping; tissue typing; forensic biology; transgenic; gene therapy; antianorectic; human; chromosome 2.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Peptide 1..32
FT Protein /label= Signal_peptide
FT Protein 33..397
FT Protein /note= "Human mature LIG46 protein"
XX WO200274905-A2.
XX 26-SEP-2002.
XX 20-NOV-2001; 2001WO-US43345.
XX 21-NOV-2000; 2000US-0717778.
XX (MILL-) MILLENNIUM PHARM INC.
XX White DW, Zhou J, Tartaglia LA, Stricker-Krongrad A, Clausen R;
XX WPI: 2002-759886/82.
XX N-PSDB; AAD46673.
XX New isolated nucleic acid molecules, designated as leptin-induced genes 46 (LIG46), useful for treating a subject having a disorder characterized by undesirable level of LIG46 expression or activity, such as low body weight.
XX Claim 22; Fig 6; 90pp; English.
XX The invention relates to LIG46, a gene whose expression is induced by leptin. LIG46 DNA and protein are useful in treating a subject having a disorder characterized by undesirable level of LIG46 expression or activity, such as low body weight. They are also useful in a screening assay, chromosomal mapping, tissue typing and forensic biology. The probes based on the LIG46 nucleotide sequence are useful for detecting transcripts or genomic sequences encoding the same or related proteins. The LIG46 polypeptides are useful as immunogens for raising anti-LIG46 antibodies. The host cells are useful for producing non-human transgenic animals. LIG46 DNA is used in gene therapy. The present sequence is human LIG46 protein. LIG46 gene is located at chromosome 2.
XX Sequence 397 AA:
Query Match 87.5%; Score 1862; DB 23; Length 397;
Best Local Similarity 86.9%; Pred. No. 1.9e-195;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;
QY 1 MSVGRRRVKKLGLMMANVFYILIVEVSKNSODKNGKGVIIIPKEFKWPPSTPAYWN 60
DB 1 MSVGRRRVKKLGLMMANVFYILIVEVSKNSODKNGKGVIIIPKEFKWPPSTPAYWN 60
QY 61 REQEKLNWYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRKDFL 120
DB 61 REQEKLNWYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRKDFL 120

QY 121 LYLRCRNYSLIDOPKKCAKPFLLAIAKSLIPHFARQAIRESWGRETNGVQTYYRVE 160
DB 121 LYLRCRNYSLIDOPKKCAKPFLLAIAKSLIPHFARQAIRESWGQESNAGNQTYYRVF 180
QY 181 LGKTPPEDNHPDLSMDLKFESDKHODILMNNYRDTFFNLSKEVFLRWVSTSCPDADF 240
DB 181 LGKTPPEDNHPDLSMDLKFESDKHODILMNNYRDTFFNLSKEVFLRWVSTSCPDADF 240
QY 241 VFKGDDVFNTHHILNLSKSKAKOLFIDGVTHNAGPHRDKKLYIPEVFTYGVY 300
DB 241 VFKGDDVFNTHHILNLSKSKAKOLFIDGVTHNAGPHRDKKLYIPEVFTYGVY 300
QY 301 PPVAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGFRTEDEE 360
DB 301 PPVAGGGFLYSGHLLALRLYHITDVHLYPIDDVYTGMCLOKGLVPEKHKGFRTEDEE 360
QY 361 KNKNKICSYIDLMLVHSRKPQEMIDINSQLOSPNLKC 397
DB 361 KNKNKICSYVDLMLVHSRKPQEMIDINSQLOSAHLKC 397

RESULT 12
ABB82649
ID ABB82649 standard; protein: 397 AA.
XX AC ABB82649;
XX AC ABB82649;
DT 19-FEB-2003 (first entry)
XX Human beta3GnT polypeptide sequence.
XX Beta3GnT; transgenic; beta-1,3-N-acetylglucosaminyltransferase;
KW disease system; human; enzyme.
XX Homo sapiens.
XX WO200279413-A2.
XX 10-OCT-2002.
XX 29-MAR-2002; 2002WO-US09645.
XX 29-MAR-2001; 2001US-280706P.
PR 28-MAR-2002; 2002US-0280706.
XX (DELTA) DELTAGEN INC.
XX Leviten MW, Phillips R;
XX WPI: 2003-067437/06.
DR N-PSDB: ABV75082.
XX New transgenic mouse comprising a disruption in a beta3GnT gene, as in
PT vivo model to study various disease states or conditions in which
PT beta3GnT may be implicated or involved, such as abnormal cell growth,
PT cancer and metastasis.
XX Disclosure; Fig 4: 55pp; English.
XX The invention relates to a transgenic mouse comprising a disruption in a
CC beta3GnT (beta-1,3-N-acetylglucosaminyltransferase) gene, where there is
CC no native expression of beta3GnT gene. The transgenic mice may be used
CC as in vivo model to study various disease states or conditions in which
CC beta3GnT may be implicated or involved, such as abnormal cell growth,
CC cancer and metastasis, and to evaluate various treatments or to identify
CC agents for treating disease states or conditions, such as anxiety or
CC depression. Animal-based disease systems may be used to identify
CC compounds capable of ameliorating disease symptoms, as test substrates
CC for the identification of drugs, pharmaceuticals, therapies and
CC interventions that may be effective in treating a disease or other
CC phenotypic characteristic of the animal. The present sequence represents
CC a human beta3GnT polypeptide sequence.

SQ Sequence 397 AA;
Query Match 87.5%; Score 1862; DB 24; Length 397;
Best Local Similarity 86.9%; Pred. No. 1.9e-195;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;
QY 1 MSVGRRRVKLLGILMANVFLYLIVEVSKNSODKNGKGVIIIPKEKFWKPPSTPRAYWN 60
DB 1 MSVGRRRRIKLLGILMANVFIYIMEVSKSSQKNGKGEVILPKEKFKISTPEAYWN 60
QY 61 REQEKLNRYNPIILNRVANOTGELATSPNTSHLSYCEPDSVTMTAVTFDNNLPDRFKDFL 120
DB 61 REQEKLNRYNPIILSMLTNOTGEAGRLSNISHLAYCEPDLRVTSVVTGFNNLPDRFKDFL 120
QY 121 LYLRCRNYSLIDOPKKCAKPFLLAIAKSLIPHFARQAIRESWGRETNGVQTYYRVF 180
DB 121 LYLRCRNYSLIDOPKKCAKPFLLAIAKSLIPHFARQAIRESWGQESNAGNQTYYRVF 180
QY 181 LGKTPPEDNHPDLSMDLKFESDKHODILMNNYRDTFFNLSKEVFLRWVSTSCPDADF 240
DB 181 LGKTPPEDNHPDLSMDLKFESDKHODILMNNYRDTFFNLSKEVFLRWVSTSCPDADF 240
QY 241 VFKGDDVFNTHHILNLSKSKAKOLFIDGVTHNAGPHRDKKLYIPEVFTYGVY 300
DB 241 VFKGDDVFNTHHILNLSKSKAKOLFIDGVTHNAGPHRDKKLYIPEVFTYGVY 300
QY 301 PPVAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGFRTEDEE 360
DB 301 PPVAGGGFLYSGHLLALRLYHITDVHLYPIDDVYTGMCLOKGLVPEKHKGFRTEDEE 360
QY 361 KNKNKICSYIDLMLVHSRKPQEMIDINSQLOSPNLKC 397
DB 361 KNKNKICSYVDLMLVHSRKPQEMIDINSQLOSAHLKC 397

RESULT 13
AAB49748
ID AAB49748 standard; protein: 413 AA.
XX AC AAB49748;
XX AC AAB49748;
DT 17-APR-2001 (first entry)
XX Human beta 1,3-N-acetylglucosamine transferase protein G3.
DE Human beta 1,3-N-acetylglucosamine transferase protein G3.
XX Sugar chain synthesis agent; beta 1,3-N-acetylglucosamine transferase;
KW inflammation; cancer; metastasis; human.
XX Homo sapiens.
XX WO200100848-A1.
XX 04-JAN-2001.
XX 29-JUN-2000; 2000WO-JP04304.
PR 29-JUN-1999; 99JP-0183437.
PR 16-MAR-2000; 2000JP-0074757.
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX Sasaki K, Shiraishi N, Natsume A, Yamada Y, Nakagawa S, Sekine S;
XX WPI: 2001-102895/11.
DR N-PSDB: AAF29255.
XX New polypeptide having beta1,3-N-acetylglucosamine transferase activity
PT for diagnosis of inflammation, cancer and cancer metastasis,
PT development of remedies, and for producing glycoconjugates.
XX Claim 1: Page 147-150; 195pp; Japanese.
XX This invention relates to a sugar chain synthesizing agent that contains

CC a polypeptide as the active ingredient, where the polypeptide has beta
 CC 1,3-N-acetylglucosamine transferase activity. The polypeptides, DNA
 CC sequences encoding them and antibodies directed against the proteins are
 CC useful in the diagnosis of inflammation, cancer and its metastasis,
 CC development of remedies, and for producing sugar chains and
 CC glycoconjugates. Sequences AAB49748 - AAB49751 represent polypeptides of
 CC the invention, having beta 1,3-N-acetylglucosamine transferase activity,
 CC and polynucleotides AAF29255 - AAF29258 represent cDNA encoding them. PCR
 CC primers used in the isolation and characterization of the cDNA sequences
 CC are represented by sequences AAF29259 - AAF29290.

XX Sequence 413 AA:

Query Match 86.6%; Score 1844; DB 22; Length 413;
 Best Local Similarity 83.5%; Pred. No. 1.9e-193;
 Matches 345; Conservative 19; Mismatches 33; Indels 16; Gaps 1;
 QY 1 MSVGRPRVKLLGILMM-----ANVFYIIVEVSKNSQDKNGKGVIIIP 44
 DB 1 MSVGRPRVKLLGILMMANVFYIIMEVSKNSQANVFYIIMEVSKNSQDKNGKGVIIIP 60
 QY 45 KEFKPKPSTPRAYNNRQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMT 104
 DB 61 KEFKWKISTPPAYNNRQEKLNRYNPILNMLTNOTGAGRLSNISHLNYCEPDLRVIS 120
 QY 105 AVTDFNLPDRKDFLLVLCRNYSLIIDQPKKCAKYPFLLATKSLIPHARQAIRES 164
 DB 121 VTGFNLPDRKDFLLVLCRNYSLIIDQPKKCAKYPFLLATKSLIPHARQAIRES 180
 QY 165 WQRETNVGNQTVRVVFLGKTPEDNHPDLSMDLKFESDKHODILMMNYRDTFNLSLKE 224
 DB 181 WQESNAGNQTVRVVFLGKTPEDNHPDLSMDLKFESDKHODILMMNYRDTFNLSLKE 240
 QY 225 VLFLRWVSTSCPDAEFEKGGDOVFVNTHHILNLSLSKAKDLFTGDIHNAHPHD 284
 DB 241 VLFLRWVSTSCPDTEFVFGKDDVFVNTHHILNLSLSKAKDLFTGDIHNAHPHD 300
 QY 285 KKLKYYIPVYGYVYPPYAGGGFLYSGPALLRLYSATSRVHLYPIDDDVYTGMCJQKLG 344
 DB 301 KKLKYYIPVYGYVYPPYAGGGFLYSGHALLRLYHITDQVHLYPIDDDVYTGMCJQKLG 360
 QY 345 LVPEKHKGRTFDIEKNKKNICSVD:MLVHSRKPQEMIDWSOLOSPNLKC 397
 DB 361 LVPEKHKGRTFDIEKNKKNICSVD:MLVHSRKPQEMIDWSOLOSAHLKC 413

RESULT 14

AAV69698

ID AAV69698 standard; Protein: 397 AA.

AC AAV69698:

XX 08-MAY-2000 (first entry)

DT Human Brainiac protein.

DE Human Brainiac protein.

XX Brainiac; human; mammalian; expressed sequence tag; EST;

KW Drosophila melanogaster; Egghead; Notch; epithelial cell

KW Fringe family; cell viability; growth regulation; cell fate; cancer;

KW psoriasis; skin lesion; nervous system disorder; developmental syndrome;

KW transgenic animal; drug screening.

XX Homo sapiens.

OS Homo sapiens.

XX Key

FT Misc-difference 104

FT /note- "Encoded by NNNNNN"

XX CA2255109-A1.

XX 17-JUN-1999.

XX 17-DEC-1998; 98CA-2255109.

XX 17-DEC-1997; 97CA-2225126.
 XX (HSCR-) HSC RES & DEV LP.
 PA Egan SE;
 PI WPI; 2000-148082/14.
 DR N-PSDB; AAZ87186.
 XX New nucleic acids encoding a murine and human Brainiac protein, useful
 PT for detecting somatic or germline DNA lesions which are responsible for
 PT developmental syndromes or diseases including cancer -
 PS Claim 13; Page31; 40pp; English.
 XX This sequence represents human Brainiac protein. The complete murine
 CC Brainiac cDNA sequence (AAZ87185) was used to screen an EST (expressed
 CC sequence tag) database to identify human Brainiac cDNA clones.
 CC which were assembled to form the human Brainiac cDNA of the present
 CC invention. Human Brainiac has significant similarity to Drosophila
 CC melanogaster Brainiac. Drosophila Brainiac and Egghead proteins regulate
 CC adhesion between epithelial cells, this activity being dependent on the
 CC presence of Notch. Drosophila Brainiac is a secreted protein, and has
 CC sequence similarities with the Drosophila Fringe proteins. Brainiac/
 CC Egghead-mediated epithelial cell adhesion is required for cell
 CC viability, cell growth regulation and cell fate specification. Wild-type
 CC or mutant forms of mammalian Brainiac proteins may therefore be used to
 CC alter epithelial cell adhesion in a mammal. Mammalian Brainiac proteins,
 CC active fragments analogues, and nucleic acids may be used to treat
 CC diseases such as cancer, psoriasis and other skin lesions, and nervous
 CC system disorders. Mammalian Brainiac nucleic acids may also be used to
 CC detect somatic or germline DNA lesions which are responsible for
 CC developmental syndromes or diseases including cancer. The mammalian
 CC Brainiac proteins and fragments or its analogues are useful as antigens
 CC in immunoassays including enzyme-linked immunosorbent assays (ELISA),
 CC radioimmunoassays (RIA) and other non-enzyme linked antibody binding
 CC assays. Non-human transgenic animals comprising nucleotide sequences
 CC encoding human Brainiac protein (AAV69698) can be used as animal models
 CC for the study of mammalian Brainiac gene function, for the screening of
 CC candidate compounds and for the evaluation of potential therapeutic
 CC interventions.
 XX Sequence 397 AA:
 SQ
 Query Match 83.8%; Score 1784; DB 21; Length 397;
 Best Local Similarity 85.6%; Pred. No. 7.3e-187;
 Matches 332; Conservative 17; Mismatches 39; Indels 0; Gaps 0;
 QY 10 LIGILMMANVFYIIVEVSKNSQDKNGKGVIIIPKFKWKPSTPRAYNNRQEKLNRW 69
 DB 10 LIGILMMANVFYIIMEVSKNSQDKNGKGVIIIPKFKWKPSTPRAYNNRQEKLNRW 69
 QY 70 YNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNLPDRKDFLLVLCRNY 125
 DB 70 YNPILNMLTNOTGAGRLSNISHLNYXXXXXXVVTGNNLPDRKDFLLVLCRNY 129
 QY 130 LLIDQPKKCAKYPFLLATKSLIPHARQAIRESWGRNVTGTVRVVFLGKTPED 189
 DB 130 LLIDQPKKCAKYPFLLATKSLIPHARQAIRESWGRNVTGTVRVVFLGKTPED 189
 QY 190 NHPDLSMDLKFESDKHODILMMNYRDTFNLSLKEVLFLRWVSTSCPDAEFEKGGDDVF 249
 DB 190 NHPDLSMDLKFESDKHODILMMNYRDTFNLSLKEVLFLRWVSTSCPDTEFVFGKDDVF 249
 QY 250 VNTHHILNLSLSKAKDLFTGDIHNAHPHDKRLKYYIPEVYTYPPYAGGGF 309
 DB 250 VNTHHILNLSLSKAKDLFTGDIHNAHPHDKRLKYYIPEVYTYPPYAGGGF 309
 QY 310 LYSGPALLRLYSATSRVHLYPIDDDVYTGMCJQKLGVLPEKHKGRTFDIEKNKKNICS 369
 DB 310 LYSGLHALRLYHITDQVHLYPIDDDVYTGMCJQKLGVLPEKHKGRTFDIEKNKKNICS 369

QY 370 IDMLVHSRKPEMIDIWSQLQSPNLKC 397
: : : : : : : : : : : : : : : : : : :
Dd 370 VDMLVHSRKPEMIDIWSQLQSAHLKC 397

RESULT 15

RESULT IS
AAB93536

AA393530
ID AA393536 standard: protein: 373 AA.

AA
AC
AAB93536:

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12898.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW

OS Homo sapiens.

XX
PN
EP1074617-A2.

XX
PD 07-FEB-2001XX
PF 28-JUL-2000: 2000EP-0116126

XX 29-III-1999. 99.TP-0248036

PR 27-AUG-1999: 99JP-0300253.
PR 11-IAN-2000: 2000IB-0118776

PR 02-MAY-2000; 2000JP-0183767.
PB 08-JUN-2000; 2000JP-0241888

XX
DA (UET T -) UET Y REC FNOT

XX

PI Ishii S, Sugiyama T,

DR WPI; 2001-318749/34.

Primer sets for synt

PT and/or diagnosis of

XX
PS Claim 8; SEQ ID 12898; 2537pp + CD ROM; English.

CC The present invention describes primer sets for

comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

AA	Sequence	373 AA;
SQ		

Query Match 82.3% Score 1753: DB 22: Length 373:

Query match 82.3%; SCORE 1733; DB 227;
Best Local Similarity 86.6%; Pred. No. 1.7e-183;

Best local similarity 80.0%; Pled. NO. 1; 7E 103;
Matches 323; Conservative 18; Mismatches 32; Indels 0; Gaps 0;

Qy	25	VEVSKNSSODKNGKNGGVIIPEKEFKWKPSTPRAVYNREQEKLNRWYNPIINRVANOTGEL	84
Db	1	MEYKSSSOEKKNGGEVIIPEKEFKWKISTPEAYWNRQEKLNRQYNPIILSMILNOTGEA	60
Qy	85	ATSPNTSHLSYCEPOSTVYMTVDENNLPDRFKDFLLYLRCRNYSLIIDOPKKCAKKPEL	144
Db	61	GRUSNIISHUNYCEPOLRVTSVVTGNNLPDRFKDFLLYLRCRNYSLIIDOPDKCAKKPEL	120
Qy	145	LLAIKSLIPHFARRQAIRESGRETNVGNQTVVRVFLLGKTPPDNDHPDLSMLKFESDK	204
Db	121	LLAIKSLIPHFARRQAIRESGOESNAGNQTVVRVFLLGQTPPDNDHPDLSMLKFESDK	180
Qy	205	HQDILMWNYRDTFFNLSEKVELFRLWYSTCQPDRAEFVKGGDDOVFVYVTHHILNLSLSK	264
Db	181	HQDILMWNYRDTFFNLSEKVELFLRWYSTCQPDTEFVKGGDDOVFVYVTHHILNLSLSK	240
Qy	265	SKAKDLFIGDVTHNAGPHRDKKLKYYIPEVFPYTCVPPYAGGGGFLYSGPALLURLYLSATS	324
Db	241	TKAKDLFIGDVTHNAGPHRDKKLKYYIPEVYVSGLYPPYAGGGGFLYSGHALURLYHITD	300
Qy	325	RVHLYPIDDVYTCMLQKLGLVPEKHKGFRTFIEEKNKNKICSYIDLMLVHSRKKPQEMI	384
Db	301	QVHLYPIDDVYTCMLQKLGLVPEKHKGFRTFIEEKNKNKICSYVYDLMLVHSRKKPQEMI	360
Qy	385	DIWSQLOSPNLKC	397
Db	361	DIWSQLOSAHLKC	373

Search completed: October 20, 2003, 23:01:58

Job time : 40.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 23:00:41 : Search time 12.5 Seconds
(without alignments)
1343.793 Million cell updates/sec

Title: US-09-804-357b-14

Perfect score: 2123
Sequence: 1 MSVGRRIKLLGILMANVF.....RKPOEMDIWSQLQSAHLKC 397

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	616.5	29.0	378	4	US-09-482-180A-2
2	557	26.2	397	4	US-09-459-133-2
3	549	25.9	389	4	US-09-459-133-13
4	479.5	22.6	325	2	US-09-055-097-6
5	328	15.4	325	2	US-09-055-097-5
6	309.5	14.6	331	4	US-09-996-243-209
7	280	13.2	378	1	US-09-055-097-1
8	140.5	6.6	372	1	US-08-207-904-10
9	94.5	4.5	309	4	US-09-134-001C-5667
10	94	4.4	406	4	US-09-134-001C-3544
11	92.5	4.4	523	1	US-08-073-383-2
12	92.5	4.4	523	5	PCT-US94-06365-2
13	91.5	4.3	523	1	US-08-428-415-2
14	91.5	4.3	523	1	US-08-379-685-2
15	91.5	4.3	523	2	US-08-854-029-2
16	91.5	4.3	523	3	US-08-428-762-2
17	90	4.2	350	4	US-09-484-035A-11
18	90	4.2	363	4	US-09-464-035A-1
19	90	4.2	363	4	US-09-849-562A-1
20	90	4.2	363	4	US-09-849-031A-1
21	89.5	4.2	354	4	US-09-574-377-24
22	87	4.1	354	4	US-09-574-377-32
23	85.5	4.0	520	4	US-09-752-165-1
24	85	4.0	624	2	US-08-756-317-9
25	84.5	4.0	734	4	US-09-585-858-9
26	83	3.9	354	4	US-09-574-377-25
27	83	3.9	707	4	US-09-228-986-80

28	83	3.9	1077	4	US-09-412-210-1	Sequence 1, Appli
29	82.5	3.9	912	2	US-08-951-871-2	Sequence 2, Appli
30	82.5	3.9	3056	1	US-08-508-836A-8	Sequence 8, Appli
31	82.5	3.9	3056	2	US-08-629-001A-3	Sequence 3, Appli
32	82.5	3.9	3056	2	US-08-874-266-2	Sequence 3, Appli
33	82.5	3.9	3056	3	US-08-642-274D-3	Sequence 3, Appli
34	82.5	3.9	3056	3	US-08-952-127-3	Sequence 3, Appli
35	82.5	3.9	3056	3	US-08-952-014C-3	Sequence 3, Appli
36	82.5	3.9	3056	4	US-09-360-416-2	Sequence 2, Appli
37	82	3.9	321	2	US-08-780-370A-2	Sequence 2, Appli
38	82	3.9	321	4	US-09-251-330-2	Sequence 2, Appli
39	82	3.9	363	4	US-09-464-035A-3	Sequence 3, Appli
40	82	3.9	663	4	US-09-641-638-653	Sequence 653, App
41	82	3.9	1226	2	US-08-540-804-12	Sequence 12, Appl
42	82	3.9	1226	2	US-08-218-265-12	Sequence 12, Appl
43	82	3.9	1226	3	US-08-521-872-12	Sequence 12, Appl
44	82	3.9	1226	3	US-08-590-399-12	Sequence 12, Appl
45	81.5	3.8	363	4	US-09-464-035A-5	Sequence 5, Appli

ALIGNMENTS

RESULT :
US-09-482-180A-2
: Sequence 2, Application US/09482180A
: Patent No. 6361985
: GENERAL INFORMATION:
: APPLICANT: Conklin, Darrell C.
: APPLICANT: Yamamoto, Gayle
: APPLICANT: Gao, Zeren
: APPLICANT: Whitmore, Theodore E.
: APPLICANT: Jaspers, Stephen
: TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING
: FILE REFERENCE: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOG, ZN5SP6
: CURRENT APPLICATION NUMBER: US/09/482,180A
: CURRENT FILING DATE: 2000-01-12
: PRIOR APPLICATION NUMBER: 60/115,721
: PRIOR FILING DATE: 1999-01-12
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: FastSeq for Windows version 3.0
: SEQ ID NO 2
: LENGTH: 378
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-482-180A-2

Query Match	29.0%	Score 616.5	DB 4	Length 378
Best Local Similarity	43.2%	Pred. No. 4.2e-60		
Matches 133	Conservative 49	Mismatches 107	Indels 19	Gaps 8
QY	96	CEPDLRVTGVTGTFNNLPDRFKDFLLYLRCRWYSLIDQPKCAKKPFLLLAIKSLTPHF	155	
Db	76	CPNHTVSSASL---SLPSRHRFLTYRCHRFNSILL-EPSCSKDTELLAIAKSQPHV	131	
QY	156	ARQAIRESWGQESN-AGNQTIVRVFLL---GQTPPENHPDLSMLKFESEKHODILMW	211	
Db	132	ERRAAIRSTWGRVGGWARGQLKLVLLGVAGSAPP-----AQLLAYESREFDDILQW	184	
QY	212	NYRDTFFNLSEKLVFLRWVSTSCPDTFVFVGDDDFVNTHHILNLYNLSLTKAKDLF	271	
Db	185	DETEDEFNLTKEHLKLRWVVAACQAHFMLGDDDFVHVPNVLEFDDGWD--FAQDLL	242	
QY	272	IGDVTHNAGPHRDKKLYIPEVYVSGL-YPPYAGGGFLYSGHLALRYHITDQVHLYP	330	
Db	243	VGDTVQALPNTNKKYFIPPSMVRATHYPYAGGGVMSRATVRRLOAIMEDAELS	302	
QY	331	IDDVYTCMLQKLGVLPEKHKGFRTFDIEEK-NKNNICSYVDLMVLVHRSKPKQEMIDWSQ	389	
Db	303	IDDVFGVCMRLRLGLSPMHAGFKTIGIRRLDPLDPLCYRLGLLVHRLSPLEMMTNAL	362	
QY	390	LQSAHLKC	397	

APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C13
CURRENT APPLICATION NUMBER: US/09/996,243
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24

7	PRIOR APPLICATION NUMBER: 60/090540
7	PRIOR FILING DATE: 1998-06-24
7	PRIOR APPLICATION NUMBER: 60/090542
7	PRIOR FILING DATE: 1998-06-24
7	PRIOR APPLICATION NUMBER: 60/090557
7	PRIOR FILING DATE: 1998-06-24
7	PRIOR APPLICATION NUMBER: 60/090676
7	PRIOR FILING DATE: 1998-06-25
7	PRIOR APPLICATION NUMBER: 60/090678
7	PRIOR FILING DATE: 1998-06-25
7	PRIOR APPLICATION NUMBER: 60/090690
7	PRIOR FILING DATE: 1998-06-25
7	PRIOR APPLICATION NUMBER: 60/090694
7	PRIOR FILING DATE: 1998-06-25
7	PRIOR APPLICATION NUMBER: 60/090695
7	PRIOR FILING DATE: 1998-06-25
7	PRIOR APPLICATION NUMBER: 60/090596
7	PRIOR FILING DATE: 1998-06-25
7	PRIOR APPLICATION NUMBER: 60/090862
7	PRIOR FILING DATE: 1998-06-26
7	PRIOR APPLICATION NUMBER: 60/090863
7	PRIOR FILING DATE: 1998-06-26
7	PRIOR APPLICATION NUMBER: 60/091360
7	PRIOR FILING DATE: 1998-07-01
7	PRIOR APPLICATION NUMBER: 60/091478
7	PRIOR FILING DATE: 1998-07-02
7	PRIOR APPLICATION NUMBER: 60/091544
7	PRIOR FILING DATE: 1998-07-01
7	PRIOR APPLICATION NUMBER: 60/091519
7	PRIOR FILING DATE: 1998-07-02
7	PRIOR APPLICATION NUMBER: 60/091626
7	PRIOR FILING DATE: 1998-07-02
7	PRIOR APPLICATION NUMBER: 60/091633
7	PRIOR FILING DATE: 1998-07-02
7	PRIOR APPLICATION NUMBER: 60/091978
7	PRIOR FILING DATE: 1998-07-07
7	PRIOR APPLICATION NUMBER: 60/091982
7	PRIOR FILING DATE: 1998-07-07
7	PRIOR APPLICATION NUMBER: 60/092182
7	PRIOR FILING DATE: 1998-07-09

Query Match	14.6%	Score	309.5;	DB	4;	Length	331;
Best Local Similarity	29.5%;	Pred.	No. 6.5e-26;				
Matches	82;	Conservative	56;	Mismatches	123;	Indels	15;
Gaps	8;						
QY	126	RNYSLIDQDKCA-KKPFLLLAIKSLTPTFARRQAIRSWGQESNAGNQTVRVFLGQ	184				
Db	61	ODFHPTLRHSNCSHQNPFLVILVTSHPSDVKAKQAIRVTWCEKKSWMGYEVLTFELLGQ	120				
QY	185	-TPPEDNHPDLSMLKPESEKHODILMWNYPDFNLSLKEVLFYRWVSTSCPTETVFVK	243				
Db	121	EAEKEDKMLAS--LEDEHLIGDIHQDFLDITYNLLTKTIMAFRWTEFCPNAKYVYMK	178				
QY	244	GDDVYVYVNTTHILNLYNSLSKTKAKDLFIGDVYHNAGPRDKKLKYYIPEVYVS-GUYPY	302				
Db	179	ITDVFINTGNLVYL--LNLNHSKFPFTGYPLIDNYSYRGFYQKTHISYOEYFPKVPFP	236				
QY	303	VAGGGGFLYSGHLRALRYHTDQVHLYPIDDVYTGMCLOKIGL--VPEKHKGRTDIE	359				
Db	237	YCSGLGYTMSRDLPVRIYEMGHVKKPFEDVTYVIGICLLNKLKVNTHIPEDNFLFLVRIH	296				
QY	360	EKNKNKNICSYVDLMLVHSRKPQEMIDIWSQLSAHLKC	397				
Db	297	----LDVQLRRVIAAHSFGSKETITFW-OVMLRNNTTC	329				

RESULT 7
US-09-055-097-1
: Sequence 1, Application US/09055097
: Patent No. 5955282
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Guegler, Karl J.

APPLICANT: Corley, Neil C.
 APPLICANT: Shah, Purvi
 APPLICANT: Patterson, Chandra
 TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESS: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: California
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/055,097
 FILING DATE: Filed Herewith
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Certone, Michael C.
 REGISTRATION NUMBER: 39,132
 REFERENCE/DOCKET NUMBER: PF-0490 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 855-0555
 TELEFAX: (650) 845-4166
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 378 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: DDBJ/EMBL/GenBank
 CLONE: 1705085
 US-09-055-097-1

	Query Match	13.2%	Score 280;	DB 2:	Length 378;
	Best Local Similarity	27.0%;	Pred. No. 1.6e-22;		
	Matches #1; Conservative	46;	Mismatches 119;	Indels 54;	Gaps 9;
Qy	130 LLIDPDKCA---KKPFLLLAIKSLTHFARRAOAIRESWGSEAGNAGNTVTVRVFLGLTGP	186			
	: :				
Dd	56 LLIPNQACSGCAPPELLILVCTAPENLNGRNAIQRASMGGLREARGLRVQTFLTLIGE--	113			
	: :				
Qy	187 PEDNHPE-----DLSDMLKFESEKHODILMWNRYDTFFNLISLKEVLFRWVSTSCPOTE	239			
	: :				
Dd	114 PNAHQPVWVGSGSDIAS-----ESAAGQDIQAAPDSYRNLTITKLSGNNAAEKICPMAR	169			
	: :				
Qy	240 VFVKGDVDVFVNTHHINYL-----NSLSKTKAKDLFIKD	274			
	: :				
Dd	170 YVLKTDVVYVNPVELSELVRLGRWRGWQWERSTEQRAEQQGGGVHLSEEVPLYLYGR	229			
	: :				
Qy	275 VHNAGPHRDKKLKYIPEVYS--GLYPYAGGGGLYSGHLALRLVHYITDOVHLXI	331			
	: :				
Dd	230 VHWVRNPSPTGCRHRVSBEOWHTWGPFPFSAGTGCVLSASAVOLIILKVASRAPLPPL	289			
	: :				
Qy	332 DVIYTGMCLQKLGHPV---KHKGFTFDIEKNKNNNICSYVDMLIV-HSRKPQEMIDIW	387			
	: :				
Dd	290 EDVFGVSNRRGLAPTCCVKLAGATHYPLDR-----CCYGFLTSHRLDPWMKQAW	343			
	: :				

RESULT 8
US-08-207-904-10
; Sequence 10, Application US/08207904
; Patent No. 5477002
; GENERAL INFORMATION:
; APPLICANT: Tuttle, AnnMarie

```

: CURRENT FILING DATE: 1998-08-13
:
: PRIOR APPLICATION NUMBER: US 60/064,964
:
: PRIOR FILING DATE: 1997-11-08
:
: PRIOR APPLICATION NUMBER: US 60/055,779
:
: PRIOR FILING DATE: 1997-08-14
:
: NUMBER OF SEQ ID NOS: 5674
:
: SEQ ID NO 5667
:
: LENGTH: 309
:
: TYPE: prt
:
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-5667

```

Query Match 4.58; Score 94.5; DB 4; Length 309;
Best Local Similarity 19.9%; Pred. No. 0.059;
Matches 56; Conservative 59; Mismatches 91; Indels 75; Gaps 12;

[illegible][illegible]

```

Qy 211 --WNYRDTFFNLSLKEVFLR-----WVSTSCPDTFFVFGKDDVFVNTH-----HI 255
      :||||| :|::| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 204 IYTAGKDTFFNDMLEKIDAKNGFDDVFGWKSVS---KESIIIRNPDTILISTECKSKSDYI 260

```

2020年12月15日

Qy 256 -----LNYLNSLSKTKAKDLFIGDVIHNAGPHRDKCLK 288

[illegible]

Db 261 EMIKKRGFDKINAVKNTRIETV-DGDEVSRPGPRIDEGLK 300

US-09-134-001C-3544
; Sequence 3544, Ap
: Patent No. 638037

; FICHE NO.: 0500570
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHY

1. TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14
 ;
 NUMBER OF SEQ. IS NO. 5074

NUMBER OF SEQ ID NOS: 5674

```

; SEQ ID NO 3544
; LENGTH: 406

```

```

; LENGTH: 406
; TYPE: DDT

```

TYPE: PRT
ORGANISM: staphylococcus epidermidis

ORGANISM: STAPHYLOCOCCUS EPIDERMIDIS
US-09-734-001C-3544

OS-00100-651-3244

Query Match 4.48: Score 94: DB 4: Length 406:

Query Match	4.4%	Score 94;	SD 4;	Length 400;
Best Local Similarity	21.98;	Pred. No. 0.1;		

BEST LOCAL SAMPLING	21.36%	PREC. NO.: 0.17	
MATCHES	87;	CONSERVATIVE	57; MISMATCHES 151; INDELS 102; GAPS 25;

[illegible]

Qy 42 IIPKEKFW-KISTPPE-AYWNRE-----QEK-LNRQYNPILSMLTNQTGEA 84

— — — — —

Db 3 LIKKLKKWLKFKSTKPYIKHWVRSKRGFM DIKAQLKQIQDKGLYRELQPIQSVEKQYI - - 59

85 GRLSNISHLNYCEPDLRVTSVVTGFNNL--PDRFKDFLLYLCRNYSLIDQPDCKAKK 141

[illegible]

Db 60 -YINDQSYINFTSND-----YLGIGQLEYQPQNFLDFI-----KTYSIHLSSRLVSGN 107

QY 142 PFLLLAI-KSLTPHFARRQAIRESWGQESNAG-----NQTVVRVFLLGQTPPEDNHPDL 194

[illegible]

Db 108 SWYQLEQALISEHFNEDALIPNSGYDANLAVFNKNNVV-IF-----SDQONHASI 161
QY 195 SDMLKFESEKHQDILMMNYRDTFFNLSLKEVFLRWVSTSCPDTEFVKGDDDDVFN-TH 253
Db 162 IDGKLSGLSVIYQHLNYDOLESHLA-----RHTNPDQVKVTSVSTNGTK 211
QY 254 HILAVNLSLTKAKDLFIGDVHINAGPHRDKLKYIP-EVYVSGLYPPYAGGG----- 308
Db 212 ADINRLVHL-KORYNAILIIDASHLGN---LFEYHADIDIVTSSLSKAWAGHGLILS 267
QY 309 -----FLYSGHL-ALRLXHTIDQVHL-YPIDVYTGMLQKLGVLPE--KH 350
Db 268 SKDIKDLINKGRSLIYSSSLPSYHLVFI-QVSLQHVIEDTYR---REKLNALSEYFNH 322
QY 351 KGRFTFIEKN-----KNNIC-----SYVDLMLVH 376
Db 323 QFMELFPNQLSNTPNIKNIIVCDLSASAOAYDMLFEH 359

RESULT 11
US-08-073-383-2
: Sequence 2, Application US/08073383
: Patent No. 5443962
: GENERAL INFORMATION:
: APPLICANT: Draetta, Giulio
: APPLICANT: Cottarel, Guillaume
: APPLICANT: Damagnez, Veronique
: TITLE OF INVENTION: {fillin}Insert Title of Application*ASSAY AND REAGENTS FOR
: TITLE OF INVENTION: ANTI-PROLIFERATIVE AGENTS)
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 60 State Street, Suite 510
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII (text)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/073,383
: FILING DATE: 19930604
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Vincent, Matthew P.
: REGISTRATION NUMBER: 36,709
: REFERENCE/DOCKET NUMBER: MII-009
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 227-7400
: TELEFAX: (617) 227-5941
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 523 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-073-383-2

Query Match 4.4%; Score 92.5; DB 1; Length 523;
Best Local Similarity 22.0%; Pred. No. 0.23;
Matches 72; Conservative 31; Mismatches 110; Indels 115; Gaps 16;
QY 71 NPILSMLTNOTG-----EAGRLSNISHLNCEPDLRVTSVYTGNNLPDRFKDFLLYLR 124
Db 204 SPVTATLSDEDDGFDVLLDGNLNKEETPSCMASLWTAPLVMTTTLNDRCK----- 256
QY 125 CRNYSLLIDQPKCAKKPFLLLAIKSLTPHFPARRQAIRESWGQESNAGNQTVVVRVFLGQ 184
Db 257 -----LFDSPSLCSS-----TRSVLKRPERSQEESPSTKRRKSMGA 296
QY 185 TTPEDNHPDLSMLKFESEKHQDILMMNYRDTFFNLSLKEVFLRWVSTSCPDTEFVKG 244
Db 297 SPKESTNPE-----KAHETLHQ-----SLSL-----ASSP-----KG 323
QY 245 DDDVFVNTTHILNLSLTKAKDLFIGDVHIN-AGPHRDKKLYIPEVYVSGL----- 299
Db 324 T-----IENILNDPRDLIGDFSKGYLFHTVAGKHQD--LKYSPEINASVINGRF 372
QY 300 -----YP-PYAGGGGFLYSGHLALRLXHTIDQVHLPIDDVYTGMLQKLG 345
Db 373 ANLIKEFVIDCRYPYEYEGG-----HIKGAVNLHMEEEVED--FLIKRPI 416
QY 346 VPEKHKG-FRTFDIEKNKN--NICSIV 370

Db 257 -----LFDSPSLCSS-----TRSVLKRPERSQEESPSTKRRKSMGA 296
QY 185 TTPEDNHPDLSMLKFESEKHQDILMMNYRDTFFNLSLKEVFLRWVSTSCPDTEFVKG 244
Db 297 SPKESTNPE-----KAHETLHQ-----SLSL-----ASSP-----KG 323
QY 245 DDDVFVNTTHILNLSLTKAKDLFIGDVHIN-AGPHRDKKLYIPEVYVSGL----- 299
Db 324 T-----IENILNDPRDLIGDFSKGYLFHTVAGKHQD--LKYSPEINASVINGRF 372
QY 300 -----YP-PYAGGGGFLYSGHLALRLXHTIDQVHLPIDDVYTGMLQKLG 345
Db 373 ANLIKEFVIDCRYPYEYEGG-----HIKGAVNLHMEEEVED--FLIKRPI 416
QY 346 VPEKHKG-FRTFDIEKNKN--NICSIV 370
Db 417 VPTDGKRVIVWFHCFESSERGRPMCRYV 444

RESULT 12
PCT-US94-06365-2
: Sequence 2, Application PC/TUS9406365
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Assay and Reagents for Identifying
: TITLE OF INVENTION: Anti-proliferative Agents
: NUMBER OF SEQUENCES: 6
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII (text)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/06365
: FILING DATE: 06-JUN-1994
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/073,383
: FILING DATE: 04-JUN-1993
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 523 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
PCT-US94-06365-2

Query Match 4.4%; Score 92.5; DB 5; Length 523;
Best Local Similarity 22.0%; Pred. No. 0.23;
Matches 72; Conservative 31; Mismatches 110; Indels 115; Gaps 16;
QY 71 NPILSMLTNOTG-----EAGRLSNISHLNCEPDLRVTSVYTGNNLPDRFKDFLLYLR 124
Db 204 SPVTATLSDEDDGFDVLLDGNLNKEETPSCMASLWTAPLVMTTTLNDRCK----- 256
QY 125 CRNYSLLIDQPKCAKKPFLLLAIKSLTPHFPARRQAIRESWGQESNAGNQTVVVRVFLCQ 184
Db 257 -----LFDSPSLCSS-----TRSVLKRPERSQEESPSTKRRKSMGA 296
QY 185 TTPEDNHPDLSMLKFESEKHQDILMMNYRDTFFNLSLKEVFLRWVSTSCPDTEFVKG 244
Db 297 SPKESTNPE-----KAHETLHQ-----SLSL-----ASSP-----KG 323
QY 245 DDDVFVNTTHILNLSLTKAKDLFIGDVHIN-AGPHRDKKLYIPEVYVSGL----- 299
Db 324 T-----IENILNDPRDLIGDFSKGYLFHTVAGKHQD--LKYSPEINASVINGRF 372
QY 300 -----YP-PYAGGGGFLYSGHLALRLXHTIDQVHLPIDDVYTGMLQKLG 345
Db 373 ANLIKEFVIDCRYPYEYEGG-----HIKGAVNLHMEEEVED--FLIKRPI 416
QY 346 VPEKHKG-FRTFDIEKNKN--NICSIV 370

Search completed: October 20, 2003, 23:05:19
Job time : 13.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:54:34 ; Search time 12.5 Seconds
(without alignments)
1493.568 Million cell updates/sec

Title: US-09-804-357B-14
Perfect score: 2123
Sequence: 1 MSVGRRRIKLGLILMANVF.....RKPEMIDIWSQLQSAHLXC 397

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2123	100.0	397	1 B3G7_HUMAN	Q9y2a9 h beta-1,3-
2	601	28.3	372	1 B3G8_HUMAN	O9y2a9 h beta-1,3-
3	418	19.7	308	1 B3G5_MOUSE	O9j167 m beta-1,3-
4	363	17.1	301	1 B3G5_PANPA	Q9n294 p beta-1,3-
5	360	17.0	300	1 B3G5_GORGO	Q9n293 g beta-1,3-
6	360	17.0	310	1 B3G5_HUMAN	Q9y2c8 h beta-1,3-
7	358	16.9	297	1 B3G5_PANTR	Q9n295 p beta-1,3-
8	298.5	14.1	371	1 B3G4_MOUSE	Q9z0f0 mus musculus
9	298.5	14.1	371	1 B3G4_RAT	O88178 rattus norv
10	280	13.2	378	1 B3G4_HUMAN	O96024 homo sapien
11	110	5.2	1026	1 BGAL_STRTR	P23989 streptococ
12	104	4.9	1487	1 MDS3_YEAST	P53094 saccharomyc
13	102.5	4.8	453	1 YAW6_SCHPO	Q10181 schizosacch
14	102.5	4.8	544	1 YB72_METJA	Q38867 methanococ
15	95.5	4.5	351	1 YZ39_METJA	Q60294 methanococ
16	94.5	4.5	485	1 SYE2_THETN	Q8r7t1 thermococ
17	93	4.4	1507	1 Y056_HUMAN	P42695 homo sapien
18	91.5	4.3	523	1 MP11_HUMAN	P30304 homo sapien
19	91.5	4.3	646	1 PIXB_HUMAN	Q14155 homo sapien
20	91	4.3	374	1 LFNG_BRARE	Q8jhf2 brachydanio
21	90	4.2	279	1 ATPG_MYCGE	P7640 mycoplasma
22	89.5	4.2	354	1 PON3_HUMAN	Q15166 homo sapien
23	89.5	4.2	619	1 DP25_PYRAB	Q9v2f3 pyrococcus
24	89.5	4.2	1110	1 VGLM_INSV	Q01260 impatiens n
25	89.5	4.2	4036	1 RPL1_DUGBV	Q66431 dugbe virus
26	89	4.2	1916	1 RIF1_YEAST	P29539 saccharomyc
27	88.5	4.2	395	1 CGA2_CHICK	P33449 gallus gall
28	88.5	4.2	664	1 YMC0_YEAST	Q03722 saccharomyc
29	87.5	4.1	512	1 V40_HSVJ7	P52351 human herpe
30	87	4.1	343	1 YD57_METJA	Q58752 methanococ
31	87	4.1	626	1 HTPG_BUCBP	Q89a93 buchnera ap
32	87	4.1	785	1 FTAL_YEAST	Q01329 saccharomyc
33	86.5	4.1	2376	1 TAO3_YEAST	P40468 saccharomyc

RESULT 1

B3G7_HUMAN

ID B3G7_HUMAN STANDARD; PRI: 397 AA.

AC Q9N97; Q9NQ9; Q9NQR0; Q9NUT9;

DT 16-OCT-2001 (Rel. 40, Created)

DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Beta-1,3-galactosyltransferase 7 (EC 2.4.1.1.) (Beta-1,3-GalTase 7)

DE (Beta3Gal-T7) (B3Gal-T7) (UDP-galactose:beta-N-acetylglucosamine beta-1,3-galactosyltransferase 7) (UDP-Gal:beta-GlcNAc beta-1,3-galactosyltransferase 7) (Beta-3-Gx-T7).

GN B3GNT1 OR B3GALT7.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

CX MCBT-TaxID=9606;

FX [1]

RP SEQUENCE FROM N.A.

RA Anado M., Carneiro F., Clausen H.;

RT "Cloning and expression of two beta-1,3-galactosyltransferases: beta3gal-T5 and beta3gal-T6.;"

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).

RC TISSUE=Urinary bladder;

RA Gromova I., Gromov P., Celis J.E.;

RT "A novel member of beta-1,3-galactosyltransferase family is down regulated during bladder ICC progression.;"

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Brain, and Testis;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buotow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.;"

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [4]

RP SEQUENCE OF 25-397 FROM N.A.

RC TISSUE=Placenta;

34 85.5 4.0 289 1 SYLB_AQUAE O67646 aquifex aeo

35 85 4.0 533 1 THAS_MOUSE P36423 mus musculus

36 85 4.0 1144 1 M2AL_HUMAN P16706 homo sapien

37 84.5 4.0 696 1 SCK1_SCHPO Q50530 schizosacch

38 84.5 4.0 704 1 NEUL_PIG Q02038 sus scrofa

39 84.5 4.0 1018 1 YC14_METJA Q58611 methanococ

40 84.5 4.0 1097 1 DPOD_YEAST P15436 saccharomyc

41 84 4.0 432 1 BFRA_THEMA O33833 thermotoga

42 84 4.0 502 1 C83A_ARATH P48421 arabidopsis

43 84 4.0 1150 1 SCC3_YEAST P40541 saccharomyc

44 83.5 3.9 452 1 PIV2_ADE12 P12540 human adeno

45 83.5 3.9 518 1 RN23_HUMAN Q9hcm9 homo sapien

ALIGNMENTS

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA Tanai H., Kinata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K.,
RA Masuo Y., Kanehori K.;
RT "NEDO human cDNA sequencing project.";
RN Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RP REVIEW.
RX MEDLINE=20047730; PubMed=10580128;
RA Amado M., Almeida R., Schwientek T., Clausen H.;
RT "Identification and characterization of large galactosyltransferase
RT gene families: galactosyltransferases for all functions.";
RL Biochim. Biophys. Acta 1473:35-53(1999).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9NY97-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9NY97-2; Sequence=VSP_001791;
CC Note=NO experimental confirmation available;
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC -!- CAUTION: WAS INDICATED AS B3GAL-T6 IN SUBMITTED DNA ENTRIES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ006077; CAR91546.1;
CC EMBL; AF288208; AAF97253.1;
CC EMBL; AF288209; AAF97254.1;
CC EMBL; BC030579; AAH30579.1;
CC EMBL; BC047933; AAH47933.1;
CC EMBL; AK002009; BAA92031.1; ALT_INIT.
CC Genew; HGNC:15629; B3GNT1.
CC GO; GO:0016021; C: integral to membrane; NAS.
CC GO; GO:0008499; F: UDP-galactose beta-N-acetylglucosamine beta-1,3-
CC InterPro; IPR002659; Glyco_trans_31.
CC Pfam; PF01762; Galactosyl_T; 1.
CC Transfase; Glycosyltransferase; Glycoprotein; Transmembrane;
CC Signal-anchor; Golgi stack; Multigene family; Alternative splicing.
CC DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
CC FT DOMAIN 29 397 LUMENAL, CATALYTIC (POTENTIAL).
CC FT CARBOHYD 79 79 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 89 89 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 173 173 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 219 219 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT VARSPLIC 1 11 MSVGRRIKLL -> MVSRLV (in isoform 2).
CC FT CONFLICT 11 11 /FTID=VSP_001791.
CC FT CONFLICT 11 11 L -> LL (IN REF. 1).
CC SQ SEQUENCE 397 AA; 46022 MW; B104ECCAE26DC4AC CRC64;
Query Match 100.0%; Score 2123; DB 1; Length 397;
Best Local Similarity 100.0%; Pred. No. 8.4e-169;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSVGRRIKLLGLMMANVFYFMEVSKSSQKNGKNGKEVFIPEKFKWISTPPEAYWN 60
DB 1 MSVGRRIKLLGLMMANVFYFMEVSKSSQKNGKNGKEVFIPEKFKWISTPPEAYWN 60
QY 61 RQEKLNQYNPILSMLNTQNGEAGRLSNISHLNYPEDLRVTSVVTGFWNLDPDFKDFL 120
DB 61 RQEKLNQYNPILSMLNTQNGEAGRLSNISHLNYPEDLRVTSVVTGFWNLDPDFKDFL 120

QY 121 LYLRCRNYSLLIQDPDKCAKPPPELLLAISLTPHFARRQAIRESWGQESNAGNQTVRVF 180
DB 121 LYLRCRNYSLLIQDPDKCAKPPPELLLAISLTPHFARRQAIRESWGQESNAGNQTVRVF 180
QY 181 LLGOTPPEDNHPDLSMDLKFSEKHDILMWNRYDRTFFNLKSLKEVFLRWVSTSCPDTDF 240
DB 181 LLGOTPPEDNHPDLSMDLKFSEKHDILMWNRYDRTFFNLKSLKEVFLRWVSTSCPDTDF 240
QY 241 VFKGDDVFEVNTHHILNLSLSKTKAKOLFIDGVIHNAAGPHRDKKLKYIPEVYVSGLY 300
DB 241 VFKGDDVFEVNTHHILNLSLSKTKAKOLFIDGVIHNAAGPHRDKKLKYIPEVYVSGLY 300
QY 301 PPYAGGGFYSGLHRLRLYHITDQVHLYPDIDVYTGMCLOKGLVPEKHKGFRTDIEE 360
DB 301 PPYAGGGFYSGLHRLRLYHITDQVHLYPDIDVYTGMCLOKGLVPEKHKGFRTDIEE 360
QY 361 KNKNKNCISYVDLMLVHSRKPQEMID1WSQLQSAHLKC 397
DB 361 KNKNKNCISYVDLMLVHSRKPQEMID1WSQLQSAHLKC 397
RESULT 2
B3G8_HUMAN
ID B3G8_HUMAN STANDARD; PRT; 372 AA.
AC Q9X2A9;
DI 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-1,3-galactosyltransferase 8 (EC 2.4.1.-) (Beta-1,3-GalTase 8)
DE (Beta3Gal-T8) (B3Gal-T8) (UDP-galactose:beta-N-acetylglucosamine beta-
DE 1,3-galactosyltransferase 8) (UDP-Gal:beta-GlcNAc beta-1,3-
DE galactosyltransferase 8) (Beta-3-Gx-T8) (Core 1 extending beta-1,3-N-
DE acetylglucosaminyltransferase) (Corel-beta3GlcNAcT).
GN B3GNT3 OR B3GALT8 OR TWEM3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99173880; PubMed=10072769;
RA Yokoyama-Kobayashi M., Yamaguchi T., Sekine S., Kato S.;
RT "Cell surface of cDNAs encoding putative type II membrane proteins on the
RT secretion from a human full-length cDNA bank.";
RL Gene 228:161-167(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21332592; PubMed=11439191;
RA Yeh J., Hiraoka N., Petryniak B., Nakayama J., Ellies L.G., Rabuka D.,
RA Hindsgaul O., Marth J.D., Lowe J.B., Fukuda M.;
RT "Novel sulfated lymphocyte homing receptors and their control by a
RT core1 extension beta1,3-N-acetylglucosaminyltransferase.";
RL Cell 105:957-969(2001).
RN [3]
RP SEQUENCE FROM N.A.
RA Jensen M.A., Bennett E.P.;
RT "Cloning of a new member of the beta 1,3 Galactosyltransferase family,
RT bl.3gal-T6.";
RC Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP REVIEW.
RX MEDLINE=20047730; PubMed=10580128;
RA Amado M., Almeida R., Schwientek T., Clausen H.;
RT "Identification and characterization of large galactosyltransferase
RT gene families: galactosyltransferases for all functions.";
RL Biochim. Biophys. Acta 1473:35-53(1999).
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its


```

DE DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.1-) (Beta-1,3-GalTase 5)
DE DE (Beta3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE DE galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
GN GN B3GALT5.
OS OS Pan paniscus (Pygmy chimpanzee) (Bonobo).
OC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.
OX OX NCBI_TaxID=9597;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RA RA Liu Y., Saitou N.;
RL RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC CC -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC CC GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC CC EFFICIENT ACCEPTOR (BY SIMILARITY).
CC CC -!- PATHWAY: Glycosylation.
CC CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; AB041415; BAA94500.1;
DR DR InterPro; IPR002659; Glyco_trans_31.
DR DR Pfam; PF01762; Galactosyl_T; 1.
KW KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack; Multigene family.
FT FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
FT FT TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT FT DOMAIN 29 >301 LUMENAL, CATALYTIC (POTENTIAL).
FT FT CARBOHYD 130 130 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 231 231 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT NON_TER 301 301
FT FT SEQUENCE 301 AA; 35222 MW; E04CC70F41CB29FD CRC64;
Query Match 17.1%; Score 363; DB 1; Length 301;
Best Local Similarity 29.2%; Pred. No. 7.6e-23;
Matches 82; Conservative 66; Mismatches 119; Indels 14; Gaps 7;
QY 116 FKD-FLLYLCRCNYSLIDOPDKCAK--PFLLLAISKSLTPHARRQAIRESWGQESNAG 172
DB 31 FKESQFVYKKGDNFLK--PDTDCROTTPFLVLLVTSSSHKQLAERMAIRQTGKERTYK 87
QY 173 NQTVRVFLGQTPPEDNHPDLSMLKFESEKHODILMNNYRDTFENSLKEVLFRAVVS 232
DB 88 GKQLKTFLLGTTSSAAETKEVDQ----ESQRHGDIIQKDFLDVYVNLTKTMGIEWVH 143
QY 233 TSCPDTEFVKGGDDVFVNTHHILNLSKTKAKDLFGDVIHNGAPHRDKLKYIIP 292
DB 144 RFPQAAAFVVKTDSDMFINDVLTLL--LKNRTTRFTGFLKLNFFIRQPSKWFVS 201
QY 293 EVVYS-GLYPYAGGGFLYSGLHALRLYHITDQVHLYPDIDVYTCMCKLGLVPEKHK 351
DB 202 KSEYPMWRYPFPCSGTGVFSGDVASQVNVSKSPYIKLEDVFGVGLCLERLIRLEELH 261
QY 352 GRTFDIEKNKNKICSVDLMVLSKRPQEMIDINSQLS 392
DB 262 SQPTF-FPGGLRFVSVCFRRIVACHFIKPTRTLLDYHQALEN 301
RESULT 5
B3G5_GORGO
ID B3G5_GORGO PRT; 300 AA.
AC QN293;

```

```

DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.1-) (Beta-1,3-GalTase 5)
DE DE (Beta3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE DE galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
GN GN B3GALT5.
OS OS Gorilla gorilla gorilla (Lowland gorilla).
OC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Gorilla.
OX OX NCBI_TaxID=9595;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RA RA Liu Y., Saitou N.;
RL RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC CC -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC CC GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC CC EFFICIENT ACCEPTOR (BY SIMILARITY).
CC CC -!- PATHWAY: Glycosylation.
CC CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; AB041416; BAA94501.1;
DR DR InterPro; IPR002659; Glyco_trans_31.
DR DR Pfam; PF01762; Galactosyl_T; 1.
KW KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack; Multigene family.
FT FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
FT FT TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT FT DOMAIN 29 >300 LUMENAL, CATALYTIC (POTENTIAL).
FT FT CARBOHYD 130 130 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 231 231 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT NON_TER 300 300
FT FT SEQUENCE 300 AA; 35104 MW; IAA11692ED9F06FA CRC64;
Query Match 17.0%; Score 360; DB 1; Length 300;
Best Local Similarity 29.3%; Pred. No. 1.3e-22;
Matches 82; Conservative 64; Mismatches 120; Indels 14; Gaps 7;
QY 116 FKD-FLLYLCRCNYSLIDOPDKCAK--PFLLLAISKSLTPHARRQAIRESWGQESNAG 172
DB 31 FKESQFVYKKGDNFLK--PDTDCROTTPFLVLLVTSSSHKQLAERMAIRQTGKERTYK 87
QY 173 NQTVRVFLGQTPPEDNHPDLSMLKFESEKHODILMNNYRDTFENSLKEVLFRAVVS 232
DB 88 GKQLKTFLLGTTSSAAETKEVDQ----ESQRHGDIIQKDFLDVYVNLTKTMGIEWVH 143
QY 233 TSCPDTEFVKGGDDVFVNTHHILNLSKTKAKDLFGDVIHNGAPHRDKLKYIIP 292
DB 144 RFPQAAAFVVKTDSDMFINDVLTLL--LKNRTTRFTGFLKLNFFIRQPSKWFVS 201
QY 293 EVVYS-GLYPYAGGGFLYSGLHALRLYHITDQVHLYPDIDVYTCMCKLGLVPEKHK 351
DB 202 KSEYPMWRYPFPCSGTGVFSGDVASQVNVSKSPYIKLEDVFGVGLCLERLIRLEELH 261
QY 352 GRTFDIEKNKNKICSVDLMVLSKRPQEMIDINSQLS 391
DB 262 SQPTF-FPGGLRFVSVCFRRIVACHFIKPTRTLLDYHQAOLE 300
RESULT 6

```

```

B3G5_HUMAN
ID B3G5_HUMAN STANDARD; PRT; 310 AA.
AC Q9Y2C3; Q9NY96; Q9PIX6; Q9PLX7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
DE (Beta3Gal-T5) (B3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE galactosyltransferase 5) (Beta-3-Gx-T5).
GN B3GALT5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colorectal adenocarcinoma;
RX MEDLINE=99230269; PubMed=10212226;
RA Isshiki S., Togayachi A., Kudo T., Nishihara S., Watanabe M.,
RA Kubota T., Kitajima M., Shiraishi N., Sasaki K., Andoh T.,
RA Narimatsu H.;
RT "Cloning, expression, and characterization of a novel UDP-
RT galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
RT (beta3Gal-T5) responsible for synthesis of type 1 chain in colorectal
RT and pancreatic epithelia and tumor cells derived therefrom.";
RL J. Biol. Chem. 274:12499-12507(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99337698; PubMed=10406968;
RA Zhou D., Berger E.G., Hennes T.;
RT "Molecular cloning of a human UDP-galactose:GlcNAc beta1,3GalNAc
RT beta1,3 galactosyltransferase gene encoding an O-linked
RT core3-elongation enzyme.";
RL Eur. J. Biochem. 263:571-576(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Amado M., Carneiro F., Clausen H.;
RT "Cloning and expression of two beta-1,3-galactosyltransferases:
RT beta3Gal-T5 and beta3Gal-T6.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Toroki Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Bagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Leirach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [5]
RP SEQUENCE OF 1-298 FROM N.A.
RA Liu Y., Saitou N.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP REVIEW
RX MEDLINE=20047730; PubMed=10580128;
RA Amado M., Almeida R., Schwientek T., Clausen H.;
RT "Identification and characterization of large galactosyltransferase
RT gene families: galactosyltransferases for all functions.";
RL Biochim. Biophys. Acta 1473:35-53(1999).
CC -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC GLCNAC(BETA1,3)GALNAc STRUCTURE. CAN USE GLYCOLIPID LC3GAL AS AN
CC EFFICIENT ACCEPTOR.

```

RESULT 7

```

B3G5_PANTR          STANDARD;          PRT;    297 AA.
ID   QN295;
DT   16-OCT-2001 (Rel. 40, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DE   Beta-1,3-galactosyltransferase 5 (EC 2.4.1.1) (Beta-1,3-GalTase 5)
DE   (Beta3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE   1,3-galactosyltransferase 5) (UDP-Gal-beta-GlcNAc beta-1,3-
DE   galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
GN   B3GALT5.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.;
RA   Liu Y., Saitou N.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC   WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC   GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOPOLIPID LC3CER AS AN
CC   EFFICIENT ACCEPTOR (BY SIMILARITY).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL: AB041414; BAA94499.1;
DR   InterPro: IPR002659; Glyco_trans_31.
DR   Pfam: PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Multigene family.
FT   DOMAIN      1   7   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    8   28   SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT   FT
FT   DOMAIN      29   >297   LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD    130   130   N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174   174   N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    231   231   N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     297   297
SQ   SEQUENCE 297 AA; 34850 MW; BB7963250A637A28 CRC64;
-----
Query Match      16.98; Score 358; DB 1; Length 297;
Best Local Similarity 29.34; Pred. No. 1.9e-22;
Matches 81; Conservative 64; Mismatches 117; Indels 14; Gaps 7;

QY 116 PKD-FLYLRCRNYSLLDIDPKCAKK--PFLLAIKSLTPHARQALRESNGQESNAG 172
DB 31 FKQSFYKKDGNFLK;---PDTCDQTPPLVLLVTSRRQAERNAIRQTMGKRTYK 87
QY 173 NOTVVRVFLGQTTPEDNHPDLSDMLKFESEKHQDILMMNYRDTFFNLSLKEVLFWRVYS 232
DB 88 GKOLKTFLLGTTSSAAETKVDQ---ESQRGDIQKDFLDVYVNLTKTMGIEWVH 143
QY 233 TSCPDTEFVKGGDDVPVNTTHILNYLNSLTKAKDI.FTGDVHNAGHRDKLKYIYP 292
DB 144 RFPQAAAFVNMKIDSMDFINVDYTELL--LKKNRTRFTFTGFLKNEFFPIQPFKWFVS 201
QY 293 EVVYS-GLYPYAGGGFLYSGHLAURLYHITDQVHLYPIDDYVTCMCLQKGLVPEKH 351
DB 202 KSETPWDRIYPPFCSTGTVSGDVASQVYNVSEVPYKLEQVFGVGLCERLINIRLEE 261
QY 352 GFRTFIDEKKNKNICSYVDLMVLVHRSKPOEMIDIW 387
DB 262 SQPTF-FPGGLRSVCFRIVACHFIKPKRTLLDIW 296

```

```

RESULT 8
B3G4_MOUSE
ID   B3G4_MOUSE          STANDARD;          PRT;    371 AA.
AC   Q920F0; Q91VC1; Q920U8; Q920U9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Beta-1,3-galactosyltransferase 4 (EC 2.4.1.62) (Beta-1,3-GalTase 4)
DE   (Beta3Gal-T4) (Ganglioside galactosyltransferase) (UDP-
DE   galactose:beta-N-acetyl-galactosamine-beta-1,3-galactosyltransferase)
DE   (GAL-T2).
GN   B3GALT4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.; FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL
RP   STAGE.
RP   TISSUE=Fetal;
RX   MEDLINE=99432136; PubMed=10502288;
RA   Daniotti J.L., Martina J.A., Zurita A.R., Maccioni H.J.F.;
RT   "Mouse beta 1,3-galactosyltransferase (GAL/GMI/GD1b synthase): protein
RT   characterization, tissue expression, and developmental regulation in
RT   neural retina.";
RJ   J. Neurosci. Res. 58:318-327(1999).
RL   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SVJ;
RA   Rowen L., Qin S., Madan A., Loretz C., Hall J., James R., Dors M.,
RA   Shaffer T., Abbasi N., Ratcliffe A., Dickhoff R., Lasky S., Hood L.;
RT   "Sequence of the mouse major histocompatibility complex class II
RT   region.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
-----
CC   SEQUENCE OF 1-370 FROM N.A.
CC   STRAIN=BLG2/Msf, BFN/2Msf, C57BL/10SnJ, CAST/Ei, HMI/Msf, MSM/Msf,
CC   NJL/Msf, pgn2, and SWN/Msf;
CC   Liu Y., Kitano T., Koide T., Shirosaki T., Moriaki K., Saitou N.;
CC   *Conspicuous differences among gene genealogies of 21 nuclear genes of
CC   five Mus musculus subspecies.*;
CC   Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC   -!- FUNCTION: Involved in GMI/GD1B/GAL ganglioside biosynthesis.
CC   -!- CATALYTIC ACTIVITY: UDP-galactose + N-acetyl-D-galactosaminyl-(N-
CC   acetylneuraminy)-D-galactosyl-D-glucosyl-N-acylsphingosine = UDP
CC   + D-galactosyl-N-acetyl-D-galactosaminyl-(N-acetylneuraminy)-D-
CC   galactosyl-D-glucosyl-N-acylsphingosine.
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
CC   -!- TISSUE SPECIFICITY: Expressed in heart, brain, spleen, kidney,
CC   lung and testis.
CC   -!- DEVELOPMENTAL STAGE: First expressed at embryonic day 3.
CC   Maintained at high levels between days 4 and 7 and declines
CC   thereafter to stabilize at low levels after day 10.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
CC   EMBL: AF110520; AAC97977.1;
CC   EMBL: AF082504; AAC69622.1;
CC   EMBL: AF100956; AAC69897.1;
CC   EMBL: AB039164; BAB68688.1;
CC   EMBL: AB039165; BAB68689.1;
CC   EMBL: AB039167; BAB68691.1;
CC   EMBL: AB039166; BAB68692.1;

```

```

DR EMBL: AB039170; BAB68694.1; -
DR EMBL: AB039171; BAB68695.1; -
DR EMBL: AB039172; BAB68696.1; -
DR EMBL: AB039169; BAB68693.1; -
DR EMBL: AB039166; BAB68690.1; -
DR MGD: MGI:1859517; B3galT4.
DR InterPro: IPR002659; Glyco_trans_31.
DR Pfam: PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack; Polymorphism.
FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 26 371 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 143 143 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 248 248 W -> R (IN STRAIN PCN2).
FT VARIANT 344 344 G -> W (IN STRAIN BLG2/MSF).
SO SEQUENCE 371 AA; 41235 MW; 323A7FA56B723B3 CRC64;

Query Match 14.1%; Score 298.5; DB 1; Length 371;
Best Local Similarity 28.4%; Pred. No. 2.3e-17;
Matches 83; Conservative 49; Mismatches 115; Indels 45; Gaps 9;

QY 130 LLIDQPKC---AKKPELLLAIKSLTPHFARRQAIRESAGQESNAGNQTVRVFLGQTP 186
DQ 56 LLISNHACGSGGPPPPFLILVCTAPEHLNORNAIRASGAIAREARFVOTLFLGK-P 114
QY 187 PEDNHPDLSMLKFESEKHQDILMWNRYRDTFFNLSLKEVFLRWVSTSCPDTEFVEKGD 246
DQ 115 RROGLADLS---ESAHRDLQASFODSYRNLTKLTSLGNVKNVPCPMARYILKTD 170
QY 247 DVFNTHHILNYL-----NSLSKTKAKD-----LFTGDVHNAGPH 282
DQ 171 DVYVNPVELSELVSGQSEQKQKEAQEETATHEHRGQAVPLLYLGRVHVRPT 230
QY 283 ROKKLKYYIPEVWYS---GLYPYVAGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMC 339
DQ 231 RTPSRHHVSEELWPNWGPFPYASGTGYVLSISAVOLILKVASRAPLPLEDVFGVS 290
QY 340 LOKLGLVPE---KHKGFTFTDIEKNKNICSYYVDLMV-HSRKPOEMIDIW 387
DQ 291 ARRGGLAPTHCVKLAGATHYPLDR-----CCYKGLLTSHKVPWQMEAW 336

RESULT 9
B3G4_RAT STANDARD: PRT: 371 AA.
AC O88178;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-1,3-galactosyltransferase 4 (EC 2.4.1.62) (Beta-1,3-GalTase 4)
DE (Beta3Gal-T4) (B3Gal-T4) (Ganglioside galactosyltransferase) (UDP-
DE galactose:beta-N-acetyl-galactosamine-beta-1,3-galactosyltransferase)
DE (GAL-T2).
GN B3GALT4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL
RP STAGE.
RC TISSUE=Brain;
RX MEDLINE=97460055; PubMed=9312075;
RA Miyazaki H., Fukumoto S., Okada M., Hasegawa T., Furukawa K.,
RA Furukawa K.;
RT *Expression cloning of rat cDNA encoding UDP-galactose:GD2 beta1,3-
RT galactosyltransferase that determines the expression of
RT GD1b/GM1/GAL1*.
RL J. Biol. Chem. 272:24794-24799(1997).
CC -!- FUNCTION: Involved in GM1/GD1b/GAL ganglioside biosynthesis.
CC -!- CATALYTIC ACTIVITY: UDP-galactose + N-acetyl-D-galactosaminyl-(N-

```

```

CC acetylneuraminyl)-D-galactosyl-D-glucosyl-N-acylsphingosine = UDP
CC + D-galactosyl-N-acetyl-D-galactosaminyl-(N-acetylneuraminyl)-D-
CC galactosyl-D-glucosyl-N-acylsphingosine.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
CC -!- TISSUE SPECIFICITY: Highly expressed in thymus, spleen, kidney and
CC testis and, to a lesser extent, in brain and liver.
CC -!- DEVELOPMENTAL STAGE: In the embryonic brain, expression begins at
CC day 12 and continues until birth. Expression is maintained at low
CC levels in adult brain.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AB003478; BAA32045.1; -.
CC InterPro: IPR002659; Glyco_trans_31.
CC Pfam: PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack.
FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 26 371 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 143 143 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 371 AA; 41254 MW; 327FB76FDACF131 CRC64;

Query Match 14.1%; Score 298.5; DB 1; Length 371;
Best Local Similarity 27.7%; Pred. No. 2.3e-17;
Matches 81; Conservative 51; Mismatches 115; Indels 45; Gaps 9;

QY 130 LLIDQPKC---AKKPELLLAIKSLTPHFARRQAIRESAGQESNAGNQTVRVFLGQTP 186
DQ 56 LLIPNPQAGSGGPPPPFLILVCTAPEHLNORNAIRASGAIAREARFVOTLFLGK-P 114
QY 187 PEDNHPDLSMLKFESEKHQDILMWNRYRDTFFNLSLKEVFLRWVSTSCPDTEFVEKGD 246
DQ 115 MQQOFADLAS---ESAAQGDVLQASFODSYRNLTKLTSLGNVKNVPCPMARYILKTD 170
QY 247 DVFNTHHILNYL-----NSLSKTKAKD-----LFTGDVHNAGPH 282
DQ 171 DVYVNPVELSELVSGQSEQKQKEAQEETATVHKEHKGQAVPLLYLGRVHVRPT 230
QY 283 ROKKLKYYIPEVWYS---GLYPYVAGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMC 339
DQ 231 RTPSRHHVSEELWPNWGPFPYASGTGYVLSISAVOLILKVASRAPLPLEDVFGVS 290
QY 340 LOKLGLVPE---KHKGFTFTDIEKNKNICSYYVDLMV-HSRKPOEMIDIW 387
DQ 291 ARRGGLAPTHCVKLAGATHYPLDR-----CCYKGLLTSHKVPWQMEAW 336

RESULT 10
B3G4_HUMAN STANDARD: PRT: 378 AA.
ID B3G4_HUMAN
AC O96024;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Beta-1,3-galactosyltransferase 4 (EC 2.4.1.62) (Beta-1,3-GalTase 4)
DE (Beta3Gal-T4) (B3Gal-T4) (Ganglioside galactosyltransferase) (UDP-
DE galactose:beta-N-acetyl-galactosamine-beta-1,3-galactosyltransferase)
DE (GAL-T2) (GalT4).
GN B3GALT4 OR GALT4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```



```

DR EMBL: M63636; AAA63533.1; .
DR PIR: A49750; A49750.
DR HSP: P00722; IBLG.
DR InterPro: IPR006101; Glyco_hydro_2.
DR InterPro: IPR006102; Glyco_hydro_2ig.
DR InterPro: IPR006104; Glyco_hydro_2SB.
DR InterPro: IPR006103; Glyco_hydro_2TRIM.
DR InterPro: IPR004200; Glyco_hydro_42C.
DR InterPro: IPR004199; Glyco_hydro_42N.
DR Pfam: PF02930; Bgal_small_C; 1.
DR Pfam: PF02929; Bgal_small_N; 1.
DR Pfam: PF00703; Glyco_hydro_2; 1.
DR Pfam: PF02836; Glyco_hydro_2_C; 1.
DR Pfam: PF02837; Glyco_hydro_2_N; 1.
DR PRINTS: PR00132; GLYDRLASE2.
DR PROSITE: PS00719; GLYCOSYL_HYDROL_F2.1; 1.
DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2.2; 1.
DR Hydrolase: Glycosidase.
KW PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 458 458 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 546 546
SQ SEQUENCE 1026 AA; 117045 MW; B453AFAD68F08C98 CRC64;

Query Match
Best Local Similarity 5.2%; Score 110; DB 1; Length 1026;
Matches 76; Conservative 56; Mismatches 113; Indels 120; Gaps 20;

QY 63 QEKLNRQYNPILSKLTNQTGEAGRLSNISHLNCEPDILRVTSV-----VTGFNN----- 111
DB 45 RQSLNGKKIHYAONTNQV-----LKDFYKTEPDETDLNFINVPGHLELQGGFSGPOYVNT 99
QY 112 -LPDRFDELLYLCRNYSLLIDOPDCAKPKPELLLAIKSLTPHFAARQAIRESWGQESN 170
DB 100 QYPWDGKEFL-----RPPQVPOES---NAVASYVXFHTLNDALDKKVFISF 143
QY 171 AGNOTVVRVFLGQ-----TPEDNHPDLSMLKPESEKHQDILMMNYRDIFFNLS 221
DB 144 QGVATSIFFVWNGNFVCGYSEDSFTPEF---E*SDYL-VEGDNKLAIVYRYSTA----- 194
QY 222 LKEVLFWRVSTSCPDTEFFVKGDQDVFNTHILN--YLSLSKTKAKDLFI-GDV--- 275
DB 195 -----SWL-----EDQDFWRLYGIFRDVLYLAIPKVVODLFFVKGDYDYO 234
QY 276 -----IHNAGPHRDKKLYIPEWYSGLYPPYAGGGGLYSGHLALRLYHI----- 322
DB 235 TKAQQLDILKTGVDYEKKIKYVLS--YEGIVT--EGDASVNGDGLSVLENLKIRP 296
QY 323 --TDQVHLXPI-----DDVTGMCGLQKGLVPEKKHGFRTFDIEKNKNKNCYSYVDM 374
DB 291 WSAESPKLYDLILHVLDD-----QWVEVPVK-VGFRFRFEIKDK-----LML 332
QY 375 VHSRK 379
DB 333 LNGKR 337

RESULT 12
MDS3_YEAST
ID MDS3_YEAST STANDARD; PRT: 1487 AA.
AC F53094;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE MDS3 protein (MCK1 dosage suppressor 3).
GN MDS3 OR YGL197W OR G1307.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=97197971; PubMed=9046087;
RA Coglievina M., Klima R., Bertani I., Delneri I., Zaccaria P.,

```

```

RA Bruschi C.V.;
RT "Sequencing of a 40.5 kb fragment located on the left arm of
RI chromosome VII from Saccharomyces cerevisiae.";
RL Yeast 13:55-64(1997).
RN [2]
RP CHARACTERIZATION.
RA Li M.B., Nelgeborn L.;
RL Unpublished observations (XXX-1997).
CC !- FUNCTION: NOT KNOWN; NEGATIVE REGULATOR OF EARLY MEIOTIC GENE
CC EXPRESSION.
CC !- SIMILARITY: Contains 3 Kelch repeats.
CC !- SIMILARITY: TO YEAST YER132C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X91837; CAA62947.1; .
DR EMBL: Z72719; CAA96909.1; .
DR PIR: S62048; S62048.
DR TRANSFAC: T03484; .
DR SGD: S0003165; MDS3.
DR GO: GO:0005737; C:cytoplasm; IDA.
DR GO: GO:0030435; P:sporulation; IGI.
DR InterPro: IPR006652; Kelch_rep.
DR Pfam: PF01344; Kelch; 2.
KW Kelch repeat; Repeat; Meiosis.
FT REPEAT 171 226 KELCH 1.
FT REPEAT 234 287 KELCH 2.
FT REPEAT 371 419 KELCH 3.
SQ SEQUENCE 1487 AA; 167073 MW; 768AEFBAB796E447 CRC64;

Query Match
Best Local Similarity 4.9%; Score 104; DB 1; Length 1487;
Matches 77; Conservative 59; Mismatches 120; Indels 158; Gaps 18;

QY 19 VFVIFMEVSKSSQKNGKEVILPKFKWKISTPPEAYNNREQEKLNQYNPILSMLI 78
DB 172 LYIFGLTVSPQSGYE-----LIATNELWKLDTLTKW-----SLLS 208
QY 79 NOTGAGRLSNISHLNCEPDILRVTS--VVTGFNNLPDRFKDFLLY----- 122
DB 209 DDPQIARRFNHFMHNNENNDNRDKLIIVGGLNNMQPVKKIDIIYNISQNCWHPKTI 260
QY 123 -----LRCRNYSLLIDOPDCAKPKPELLLAIKSLTPHFAARQAIRESWG 166
DB 269 QPMEITTVNGIPLALSQDNFSLVENNE--ANVPALAF-----YMRSDQIDEVLG 318
QY 167 QESN--AGNQTIVRVFLGQTP-----PEDNHPDLSMLKPESEKHQD 207
DB 319 KDSSKIKENSPVIALPLLSSEQIGRMPSNPALPKLLNVPEYELLAPTGDYVGF-----N 372
QY 208 ILMW-----NYRDTFFNLKLEVLFLRW--VSTSCPDE-----FVEKG----- 244
DB 373 IIGGFHFNYSQSNFCHFIYDINSKWSRVRTACDIDINKHFRWRFVFWKSHHQTILLG 432
QY 245 --DDDFVFNTHILNLSLS-----KTKAKDLFIGDVIHINAGPHRDKK 286
DB 433 TKTDYDYSVSVQRFHDLSTFGFLPLNFIENKTIQLPHHKISASSLPI--PIENFAKHDT 490
QY 287 LK-----YYIPEVYVYSGL---YPPTAGGGGLYSGHLALRLY 320
DB 491 LKKVSFTSSQSFENYIRIYAPPLEMSIOSVFPFYA-----MVLGKDALEIY 539

RESULT 13
YAW6_SCHPO
ID YAW6_SCHPO STANDARD; PRT: 453 AA.
AC Q10181;

```

```

RESULT 14
ID YE72_METJA
YE72_METJA STANDARD: PRT: 544 AA.
AC Q58867;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1472 precursor.
GN MJ1472.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
GX NCBI_TaxID=2190;
[1]
SEQUENCE FROM N.A.
STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=6688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodex A.,
RA Scott J.L., Goodenough N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
*Complete genome sequence of the methanogenic archaeon, Methanococcus
Jannaschii.;
RL Science 273:1058-1073(1996).
-----
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions in no way
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: U67588; AAB99487.1; .
CC TIGR: MJ1472; .
CC
CC Hypothetical protein; Signal: Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 544 HYPOTHETICAL PROTEIN MJ1472.
SQ SEQUENCE 544 AA; 63400 MW; A3E93AD51C94EBF0 CRC64;
-----
Query Match 4.8%; Score 102.5; DB 1; Length 544;
Best Local Similarity 20.2%; Pred. No. 0.7;
Matches 61; Conservative 53; Mismatches 115; Indels 73; Gaps 15;
Qy 42 IIPKPKFKWISTPEAYNREQELKNQYNPILSMILNOTGTGAGRLSNISHUNICEPDLR 101
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
71 INKRRKFTASSEAVAY-----ITSYIKNETKES--LNNVNY-GYSNYSYN 113
Qy 102 VT-SVVTGNNLPDRKDFLLYLRCL--NYSLLIDDPDKCAKKPFLL--LAIKSLTPHF 155
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
114 ISSV-----KISPIYDPLVHLYCEIDIKYSKLLNNGELIALKPIVINRDIKLSRIPDPY 168
Qy 156 ARQAIRSEWGOES-----NAGNOTVYVFLGCTPEDNHPDLSMDLKFESEKHQDILM 210
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
169 VYLNIFYVTGYEKKININYPNDNYNRTECIILNSNFNY---SEMHNPOSTELRVIG 225
Qy 211 WNYRDTFFNLSLKEVFLRWVYSTSCPTEVFVGDDDVV-----NTHILNY----- 258
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
226 WD-----SVSKNILLPYWQV-----WRGNDVSVIWRANKNEITNNNGQYI 272
Qy 259 ---LNSLSKTKAD-----LFIGDIVHNAGPHRDKLKYIPEVYVSLGPPYAGGGGFL 310
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
273 YILNSTPTVDQDPEHTEIFDD-FNYFNPDKWDSVGVI---INNSKITVIAGGSSV 328
Qy 311 YS 312
Db : :
329 YT 330
Db : :

```

RESULT 15

```
YZ39_METJA          STANDARD;          PRT;   351 AA.
AC Q60294;
DT 01-NOV-1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJEC139.
GN MJEC139.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT *Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.*;
RL Science 273:1058-1073(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L77118; AAC37108.1; -.
DR PIR; F64514; F64514.
DR TIGR; MJEC139; -.
DR Pfam; PF04473; DUF553; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 91 111 POTENTIAL.
SQ SEQUENCE 351 AA; 41437 MW; AACBF7C32BC213B0 CRC64;

Query Match
Best Local Similarity 25.6%; Score 95.5; DB 1; Length 351;
Matches 54; Conservative 27; Mismatches 35; Indels 95; Gaps 15;

QY 216 TFFNLS-----LKEYLF---LRVYSTSCPTEFVFGDDDFVFNTHILNLSLS 263
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 TFFTLKCDKSKILKLEGLFTLSLWISTS-----LKTIL 156
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 264 KTK---AKDL--FIDVTHNAGPHDKLKYIPEVYVSGLYPPYAGGGFLYSCHLALR 318
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 157 KTKYAICRDYAKLTSAILHNL-----NIKHYF-----LVYPTHAVA 193
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 319 L-----YHITDQ-VHLYPIDDVTGMCLOKGLVPEKHGKGFRTFDIEE-----KN 362
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 194 VVIDDYVVIDOKLFIYKID-----VWLKKLG--KEKYIYTPVDIYNSKLKFEVKKYN 246
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 363 KNNICSYV--DLMLVHSRKPQEMIDWSQLQ 391
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 247 ENNLAKSEISDDIL----RKIEE--DVKKELQ 271
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: October 20, 2003, 23:02:32
Job time : 14.5 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:56:05 ; Search time 18.5 Seconds
(without alignments)
2063.728 Million cell updates/sec

Title: US-09-804-357B-14
Perfect score: 2123
Sequence: 1 MSVGRRRKLLGILMMANVF.....RKQPMIDWSQSAHLKC 397

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	11.2	739	T10648	hypothetical prote
2	215	10.1	507	T18639	hypothetical prote
3	207	9.8	345	T20031	hypothetical prote
4	194	9.1	657	G86397	protein T7N9.18 [i
5	193.5	9.1	642	D96777	hypothetical prote
6	192.5	9.1	262	T20203	hypothetical prote
7	187	8.8	357	T20029	hypothetical prote
8	187	8.8	376	T20876	hypothetical prote
9	184	8.7	384	T20446	hypothetical prote
10	179	8.4	332	T20028	hypothetical prote
11	169.5	8.0	684	F86394	hypothetical prote
12	154.5	7.3	376	T24925	protein T24P13.20
13	153.5	7.2	399	H86453	avr9 homolog F9L11
14	145	6.8	395	A86458	probasle elicitor
15	140	6.6	398	B86353	protein F2E2.6 [im
16	137.5	6.5	338	T24743	hypothetical prote
17	137.5	6.5	401	A86251	hypothetical prote
18	137	6.5	348	T31918	hypothetical prote
19	130	6.1	349	T24744	hypothetical prote
20	128.5	6.1	325	T24762	hypothetical prote
21	128.5	6.1	406	T04817	hypothetical prote
22	127.5	6.0	368	T15096	hypothetical prote
23	126.5	6.0	359	F87937	protein F14B6.6 [i
24	126.5	6.0	385	T20879	hypothetical prote
25	121	5.7	279	T20443	hypothetical prote
26	121	5.7	548	C86456	unknown protein [i
27	116.5	5.7	479	D86187	hypothetical prote
28	114.5	5.4	560	A86214	hypothetical prote
29	113.5	5.3	253	T22387	hypothetical prote

```
30 113.5 5.3 409 2 A84733 hypothetical protein T13K14.220 - Arabidopsis thaliana
31 .12 5.3 455 2 T29555 hypothetical prote
32 111.5 5.3 404 2 A86186 hypothetical prote
33 110 5.2 1026 2 A49750 beta-galactosidase
34 108 5.1 334 2 T32256 hypothetical prote
35 107.5 5.1 414 2 B96808 protein F28K19.2 [
36 107 5.0 491 2 S58330 probable membrane
37 105.5 5.0 723 2 H85092 hypothetical prote
38 105 4.9 353 2 C96573 protein F12M16.19
39 104 4.9 1487 2 S62048 probable membrane
40 102.5 4.8 453 2 T38707 probable initiator
41 99.5 4.7 544 2 G64483 hypothetical prote
42 99.5 4.7 1024 2 G72041 exodeoxyribonuclea
43 99.5 4.7 1024 2 D81624 exodeoxyribonuclea
44 99.5 4.7 1024 2 F86582 exodeoxyribonuclea
45 98.5 4.6 489 2 T13026 hypothetical prote
```

ALIGNMENTS

RESULT 1

T10648

hypothetical protein T13K14.220 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999

C:Accession: T10648

R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Le

submitted to the Protein Sequence Database, June 1999

A:Reference number: Z16991

A:Accession: T10648

A:Molecule type: DNA

A:Residues: 1-739 <BEV>

A:Cross-references: EMBL:AL080282; GSPDB:GN00062; ATSP:T13K14.220

A:Experimental source: cultivar Columbia; BAC clone T13K14

C:Genetics:

A:Gene: ATSP:T13K14.220

A:Map position: 4

A:Introns: 59/3; 86/1; 358/1; 438/3; 535/3; 578/3; 634/3; 670/3

Query Match 11.2%; Score 238; DB 2; Length 739;

Best Local Similarity 25.4%; Pred. No. 1.6e-11;

Matches 68; Conservative 56; Mismatches 108; Indels 36; Gaps 10;

Qy 142 PF-LLLAISKLTPTFARRQAIRESWGQESNAGNOTVVRVFLLGQTTPEDNHPDLSMLKF 200

Db 492 PFLFMGVLSTATNHFSEMAVRKTMQHPSIKSDVVAREFVALNP----RKEVNAMLKK 547

Qy 201 ESEKHQDILMNYRDTFFNLSLKEVLEFLRWVSTSCPDTEFVKGDDDDVFVNTTHILNLYN 260

Db 548 EAEYFGDIVILPFMDRYELVVLKTIACEFGNVTAP---YIMKDDDTFIRVESILKQID 604

Qy 261 SLSTTKAKDIFIGDVINAGPHRDKKLYK---YIPEVYSGLYPPYAGGGFYLSGHLAL 317

Db 605 GVSPEKS--LYMGNLNLRHRLPTGKWTWEEWPEAV---YPPYANGPGYIISNIA- 657

Qy 318 RLYHTITQ----VHLYPIDDVYTGMLQKLGVLPEKHKGFTFDIEEKNKNICSV--- 369

Db 658 --KIVISQNSRHKRLFLFMDVSMGLWVEQFNASMQ-----PVEYSHSWKFCQYGTCT 707

Qy 370 VDLMLVSHRKPQEMIDIWSQLQSAHLAK 397

Db 708 LNYTYAHYQSPSQMCLMDNLLKGRPQC 735

RESULT 2

T18639

hypothetical protein B0024.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T18639

R:McMurray, A.

submitted to the EMBL Data Library, April 1996

A:Reference number: Z19001

A:Accession: T18639

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-507 <WIL>

A:Cross-references: EMBL:Z71178; PIDN:CAA94876.1; GSPDB:GN00023; CESP:B0024.3

A:Experimental source: clone B0024

C:Genetics:

A:Gene: CESP:B0024.3

A:Map position: 5

A:Introns: 69/2; 118/2; 143/2; 181/1; 246/2; 281/3; 393/2; 437/3

Query Match 10.1%; Score 215; DB 2: Length 507;
Best Local Similarity 22.6%; Pred. No. 7.6e-10;
Matches 91; Conservative 73; Mismatches 156; Indels 82; Gaps 16;

Qy 58 YNRE--QEKLNQYNPILS-----
Db 125 YDSRYAQEKISPTFFVMSHLPMDLDPDYFLDLDQCLIVVECTVVIILENTYSNVTV 184
Qy 87 LSNISHLNCEPDLRTSVTVTFNNLPDRFKD-----FLLYLRGRNYSLLIDOPDK-C 138
Db 185 FLNLDIRYSYSDAPIFEVTSNFSKSHLSKOGSTNOFYHAQKDNHTYOFITVPKKOC 244
Qy 139 AKKPFLLLAIKSLTPHFARRQAIRESWGQESNA---GNQTVVVRVFLLGOTPEDNHPDLS 195
Db 245 SNTKLIQITLSTAGNFDIRQAIRETWANNPNSEHVANNDRVIRISLKSNE-----FLN 300
Qy 196 DMLKFESEKHODILMNNYKDTFNLSLKEVLEFLRWVWTSCTPTEVFVGDDSDVFVNTTH; 255
Db 301 FALQKEIEKFDMDIVDLYESVLLKLVHAILSKYKSHCOLADPQLKIDDMOAVMDMGL 360
Qy 256 LNYLSLSLTKAKDLFGIVGNAGPHROKKYKYYIPEVVIS-GLYPYVAGGGFLYSGH 314
Db 361 YRSLEDKKQASINGI-SGLIWNKSPVREKKHNVYPTLYSEKFPFPYIDGPVILIGN 419
Qy 315 LAURLYHITDQVHLYPIDVY-TGMCLOKLGLVPEKHKGFRTFDIEKN-KNNICSYVDL 372
Db 420 AVPRMLEEAKNYNQWIEDVFVTG-----IG-----KALKIKQINWANHLLRYVIE 466
Qy 373 ML-----VHSRK-POEMIDINSQSLKLC 397
Db 467 LIPSRLKCSKGGVPLIYAVHNMKGPQNIHGYQKLGK--VKC 506

RESULT 3

T20031 hypothetical protein C47F8.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T20031; T22806

R:McLay, K.

submitted to the EMBL Data Library, December 1997

A:Reference number: Z19212

A:Accession: T20031

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-345 <WIL>

A:Cross-references: EMBL:AL009246; PIDN:CAAL5841.1; GSPDB:GN00019; CESP:C47F8.3

A:Experimental source: clone C47F8

R:Kershaw, J.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19617

A:Accession: T22806

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-345 <WIL>

A:Cross-references: EMBL:Z81553; PIDN:CAB04502.1; GSPDB:GN00019; CESP:C47F8.3

A:Experimental source: clone F56H6

C:Genetics:

A:Gene: CESP:C47F8.3

A:Map position: 1

A:Introns: 23/1; 49/1; 169/3; 236/2; 278/3; 308/3

C:Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

Query Match 9.8%; Score 207; DB 2: Length 345;

Best Local Similarity 23.1%; Pred. No. 2.1e-09;

Matches 86; Conservative 72; Mismatches 130; Indels 84; Gaps 16;

Qy 48 FWKISTPPPEAYNRBEKLNQYNPILSMLTNOTGAGRSLNLSHLYNCEPDLRTSVTV 107
Db 24 FWKLPNPKSSFYLEEHQVSGWN--ISTLSRPNIDFG-----SSFIV 64

Qy 108 GFNNLPDRKDFLLYL-RCRNYSLLDQPKCAKPFLLLAIKSLTPHFARRQAIRESWG 166

Db 65 SPANIHKSOK--WYLPKKNVYTKHD-----ILMLVVSXTKNFARRNVLSTWM 112

Qy 167 QESNAGNQTIVR---VFLGOTPPEDNHPDLSMLKFESEKHODILMNNYRTFNLSLK 223

Db 113 NKENSEMKSGRMHALLFFVGLVPGDN---LAKLVLEAEIHGDMVVDLEDYDNLPPK 169

Qy 224 EVLFLRWVSTSCPDTEFVEKGGDDVFVNTHTL-----NYLNS--LS-----KTKAKDLFI 272

Db 170 TLALLIGTSKASQFIKIDDDVMFFDQDLPMLDRNFVNSNTLSYGHUSTAEELVL 229

Qy 273 GDVHNAGPHRDKKLYIYIPEVVIS-GLYPYVAGGGFLYSGHLA-LRLYHITDQVHLYP 330

Db 230 -----RNKTEPWYVPETAYNCTVYVYVMGPIYLVTKDAASLILDNANHQOFMTV 279

Qy 331 IDDVYTMCLQKLGIVPEKHKGFRTFDIEB--KNNKNICSYVDLMLVH----- 376

Db 280 EDALITGITIAQKLG-----RRYSLPNVFRNRNDITEGDQDVLAWHVQTKNDSEYKSI 331

Qy 377 -SRKPOEMIDIW 387

Db 332 FTKKLSGTSFW 343

RESULT 4

G86397

protein T7N9.18 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: G86397

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Abo

ansen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, D.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marz

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G86397

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-657 <STO>

A:Cross-references: GB:AE005172; NID:g8778858; PIDN:AAF79857.1; GSPDB:GN00141

C:Genetics:

A:Gene: T7N9.18

A:Map position: 1

Query Match 9.1%; Score 194; DB 2: Length 657;

Best Local Similarity 27.4%; Pred. No. 6e-08;

Matches 66; Conservative 43; Mismatches 106; Indels 26; Gaps 9;

Qy 158 ROATRESWGQESNAGNQTIVRVLGQTPPEDNHPDLSMLKFESEKHODILMNNYRTF 217

Db 425 RMAVRKSMQKLVSRSSKVVARFVALHARKEYNVD---LKEAEYFGDIVIVPMHY 480

Qy 218 FNLSLKEVLEFLRW-VSTSCPDTEFVEKGGDDVFVNTHTLNLNSLSKTKAKD-LFIGDV 275

Db 481 DLVVLKTVATICEYGVNTVA--AKYVMKCDODTEVRVDV---IQEAEKYKGRESLYIGNI 535

R; Dobson, R.
submitted to the EMBL Data Library, November 1996
A: Reference number: Z19276
A: Accession: T20446
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-384 <MIL>
A: Cross-references: EMBL:Z81492; PIDN: CAB04032.1; GSPDB: GNM00019; CESP: E03H4.11
A: Experimental source: clone E03H4

A: Experimental source: ClonE E03H4
C: Genetics:
A: Gene: CESP:E03H4.11
A: Map position: 1
A: Introns: 46/1; 76/1; 196/3; 263/2; 305/3; 335/3
C: Superfamily: Caenorhabditis elegans hypothetical protein T09P5.1

[illegible]

158 VGLTGDYK--MKRWQEAKLVDIIIVDMNDTYEELTKSLAILLYGVSKAPRYQMI 214
242 FKGGDDVFNTHIILNYLSLSKTKAKDLFGDVIHNGP-----HRDKKL 287

Db 215 GRIDEVI-----FFPKLTALYEQGII-DATPLCANGYKIQANGARIFDKND 261

Qy 288 KYIIEPVVYS-GLYPYAGGGFLSYGHLAURLYHITDQVHLYPIDOVY-TGMCLQKLGL 345

```

262 DB RWVPSSYSCSKFPEYVSGMLYMYTWEAAQIIKSTKYRDFIQVEDVFLTGIAEDLGI 321
346 QY VPEKHKGRTF--DIEEKNKNKICSYVDLMVLVS 377
322 DB SVRNLPKFKYKPNIDIESKS-----VDIIAHWN 349

RESULT 10

```

T20028
hypothetical protein c47P8.5 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T20028
R:McLay, K.

submitted to the EMBL Data Library, December 1997
A:Reference number: Z19212
A:Accession: T20028
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-332 (SWI)
A:Cross-references: EMBL:AL009246; PIDN:CAAL5838.1; GSPDB:GNC0019; CESP:C47F8.
A:Experimental source: clone C47F8
C:Genetics:
A:Gene: CESP:C47F8.5

A:Map position: 1
A:Introns: 26/1; 146/3; 213/2; 255/3; 285/3
C:Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1
Query Match 8.4%; Score 179; DB 2; Length 332;
Best Local Similarity 23.8%; Pred. No. 4.1e-07;
Matches 68; Conservative 55; Mismatches 123; Indels 40; Gaps 12

Qy	103	TSVTFGNFLDPREFDLYLLRCRNYSLIDPOKCAKKDFLLAIKSLTPHARQAIR	162
Db	37	TSFIISFADIQTYK-LWYL-----PÉTDIVLRSPDILAMVASRTDSFARNVLR	85
Qy	163	ESWGQESNAGNOTVVR-----VFLIGOTPEDNHDPDLSMLKPFESKHDILMWNVROT	216
Db	86	KTW---MKNYSEIIVDGRMKALFLGVW-SEDYR-VRRIVMEAKLYGMVVILDETI	139

Db 157 GDELKRLTEKGIIMRFVIGH8-----SSPGGVLDHTIEAEEQHKDFRLNHHIEGYHEL5 212
QY 222 LKEVLFL-----RWVSTSCPOTEFVFKGDDVFNTHHILNKLNS-LSKTKAKD-LFQGD 274
Db 213 SKTOIFYSAVAKW-----DADFVIKVDVDDVHN-----LGMGSTLARHRSKPRVYIGC 262
QY 275 VTHNAGP-HRDKKLYIPEVYVS-----LYPPYAGGGFLYSGHLALRLYHITDQVHL5 329
Db 263 M--KSGPVLQAGVYHEPEYKFGEGNKYFRHATGQIYAIKSKDLATYISVNRQLLHKY 320
QY 330 PDDVVTGMCLOKGLVPEKHKGFRT-----FDTEENK-----NNICSYV 370
Db 321 ANEDVSLGSNF--IGLDVE-HIDDRSLCCGPLDCEWKGAGNCPAASFDFSCSICKSV 377
QY 371 DLML-VHSRKPQEMDIW 387
Db 378 DRMLEVHORCGEGGAIW 395

RESULT 14
A86458
probable elicitor response protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
C:Accession: A86458
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.W.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A86458
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-395 <STO>
A:Cross-references: GB:AE005172; NID:g10645491; PIDN:AAG21603.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 6.8%; Score 145; DB 2: Length 395;
Best Local Similarity 27.0%; Pred. No. 0.00034;
Matches 74; Conservative 43; Mismatches 99; Indels 58; Gaps 16;

QY 114 DRKDFLLYLCRNYSLLIDQDKAKKDFLLLAISLTPHFARRQAIRE5W---GQ--- 167
Db 104 DRSEFWSESAKNSQSL-----QKFAVIGINTAFSSKKRDSVRQTMPTGEK 155
QY 168 --ESNAGNQTVVRVFLLGOTPPEDNHPDLSDMLKFESEKHODILMNYRDTFFNLSLKEV 225
Db 156 KIEKEG--IVVR-FVIGHSAPGG--VLDKAIDEDSEKDFLRKHHGHHQLSTKR 210
QY 226 LFLRWVSTSCPDTFVFKGDDVFNTHHILNKLNSLSTKAK-----DLFTGDVHINAG 280
Db 211 LYFS-TATAMYDAEFYKVDVDDVHN-----LGMVLTTLARYQSRPRIVYIGCM--KSG 260
QY 281 PHRDKK-LKYIPEVYVS-----LYPPYAGGGFLYSGHLALRLYHITDO--VHLYPTDD 333
Db 261 PVLSQKGVYHEPEYKFGEGNKYFRHATGQIYAIKSKDLA--TYISTNQGILHRYANED 318
QY 334 VYTGMCLOKGLVPEKHKGFRTDIEKNKNNIC 367
Db 319 VSLGANW--LGL-----EVEHVDERSMC 339

RESULT 15

B86353
protein F2E2.6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B86353
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alo
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzi
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall
ker, M.; Wu, D.; Yu, G.; Fraser, C.W.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: B86353
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <STO>
A:Cross-references: GB:AE005172; NID:g9280694; PIDN:AAF86563.1; GSPDB:GN00141
C:Genetics:
A:Gene: F2E2.6
A:Map position: 1

Query Match 6.6%; Score 140; DB 2: Length 398;
Best Local Similarity 25.1%; Pred. No. 0.00089;
Matches 62; Conservative 40; Mismatches 95; Indels 50; Gaps 11;

QY 141 KPFLLLAIKSLTPHFARRQAIRE5W---GQ---ESNAGNQTVVRVFLLGOTPPEDNHPDLS 195
Db 125 KVFVIGINTAFSSRRKRDSLRETWMPQGEKLEKKEKGIWVKEMIGHSSITPNSM---LD 182
QY 196 DMLKFESEKHQDILMNYRDTFFNLSLKEVFL-----RWVSTSCPDTFVFKGDDVYV 250
Db 183 KEIDSEDAQYNDFRLDHVGYINLSAKTKSFESSAVAKW-----DAEFYVKIDDDVHV 236
QY 251 NTHHILNKLNSLSTKAKD-----LFIGDVIHINAGPHRDKK-LKYIPEVYVS-----LY 300
Db 237 N-----LGTLASTLASHRSKPRVYIGCM--KSGPVLTKTKAKYREPEFVKFGESEGNKY 287
QY 301 PVYAGGGFLYSGHLALRLYHITDOVHLYPIDDVYTGMCLOKGLVPEKHKGFRTDIEE 360
Db 288 FRHATGQIYAIKSKDLATYISNQPILHRYANEDVTLG-----SWFICLEVEQ 334
QY 361 KNKNNIC 367
Db 335 IDDRNFC 341

Search completed: October 20, 2003, 23:04:47

Job time : 19.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:55:10 ; Search time 42 Seconds

(without alignments)
2439.212 Million cell updates/sec

Title: US-09-804-357b-14

Perfect score: 2123

Sequence: 1 MSVGRRRRIKLGILMANVF.....RKPEMIDWSQLSAHLKC 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rhodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacterioph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1872	88.2	397	11 Q91V18	Q91V18 mus musculus
2	1866	87.9	397	11 Q92222	Q92222 mus musculus
3	1232.5	58.1	406	13 Q8UWM0	Q8UWM0 brachydanio
4	1145.5	54.0	412	13 Q8UWM3	Q8UWM3 brachydanio
5	680.5	32.1	412	13 Q8UWM2	Q8UWM2 brachydanio
6	668	31.5	401	4 Q8NFLO	Q8NFLO homo sapien
7	660.5	31.1	397	11 Q8K437	Q8K437 mus musculus
8	651.5	30.7	397	11 Q8K0J2	Q8K0J2 mus musculus
9	651	30.7	418	13 Q8UWM4	Q8UWM4 brachydanio
10	635.5	29.9	350	11 Q923H4	Q923H4 mus musculus
11	632.5	29.8	353	4 Q8N934	Q8N934 homo sapien
12	628.5	29.6	353	4 Q96QH5	Q96QH5 homo sapien
13	628.5	29.6	353	4 Q8WY02	Q8WY02 homo sapien
14	628.5	29.6	353	4 Q8WWR5	Q8WWR5 homo sapien
15	628.5	29.6	378	4 Q8C0J1	Q8C0J1 homo sapien
16	628.5	29.6	378	4 Q8N5W4	Q8N5W4 homo sapien

17	628.5	29.6	388	4 Q8ND21	Q8ND21 homo sapien
18	616.5	29.0	384	4 Q8TDX1	Q8TDX1 homo sapien
19	615.5	29.0	390	13 Q8UWM1	Q8UWM1 brachydanio
20	603.5	28.4	377	4 Q96EK0	Q96EK0 homo sapien
21	600	28.3	372	4 Q9C0J2	Q9C0J2 homo sapien
22	573	27.0	370	11 Q8BK98	Q8BK98 mus musculus
23	573	27.0	399	11 Q8V116	Q8V116 mus musculus
24	572.5	27.0	372	11 Q8R0J2	Q8R0J2 mus musculus
25	569.5	26.8	372	11 Q9D722	Q9D722 mus musculus
26	549.5	25.9	277	4 Q8TAZ4	Q8TAZ4 homo sapien
27	549	25.9	389	11 Q8R319	Q8R319 mus musculus
28	507	23.9	374	4 Q8WWR6	Q8WWR6 homo sapien
29	479.5	22.6	326	4 Q9Y526	Q9Y526 homo sapien
30	479.5	22.6	326	6 Q9MYM7	Q9MYM7 pongo pygma
31	479.5	22.6	326	11 Q54904	Q54904 mus musculus
32	462	21.8	305	11 Q91V52	Q91V52 mus musculus
33	462	21.8	305	11 Q920V5	Q920V5 mus spicille
34	430.5	20.3	409	11 Q920V2	Q920V2 mus spicille
35	427.5	20.1	422	4 Q43825	Q43825 homo sapien
36	427	20.1	409	11 Q91V19	Q91V19 mus musculus
37	427	20.1	409	11 Q920V3	Q920V3 mus musculus
38	427	20.1	409	11 Q91V58	Q91V58 mus musculus
39	427	20.1	409	11 Q920V4	Q920V4 mus musculus
40	427	20.1	409	11 Q91VE9	Q91VE9 mus musculus
41	427	20.1	422	11 Q54905	Q54905 mus musculus
42	427	20.1	422	11 Q8CBX4	Q8CBX4 mus musculus
43	427	20.1	422	11 Q8BH19	Q8BH19 mus musculus
44	425.5	20.0	422	4 Q8BZQ9	Q8BZQ9 homo sapien
45	416.5	19.6	378	4 Q9BYG0	Q9BYG0 homo sapien

ALIGNMENTS

RESULT 1

Q91V18					
ID	Q91V18	PRELIMINARY;	PRT:	397	AA.
AC	Q91V18:				
DT	01-DEC-2001 (Tremblrel. 19, Created)				
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)				
DT	01-OCT-2002 (Tremblrel. 22, Last annotation update)				
DE	Beta-1.3-N-acetylglucosaminyltransferase (Beta-1.3-N-				
DE	acetylglucosaminyltransferase 1).				
GN	B3GNT1				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C.B17;				
RA	Egan S.E., Cohen B.L., Sarkar M., Ying Y., Cohen S., Singh N.,				
RA	Wang W., Flock G., Goh T., Schachter H.,				
RT	"Molecular cloning and expression analysis of a mouse UDP-				
RT	GlcNAc:Gal(beta1-4)Glc(NAC)-R beta1.3-N-acetylglucosaminyltransferase				
RI	homologous to Drosophila melanogaster Brainiac and the beta1,3-				
RI	galactosyltransferase family."				
RL	Glycoconj. J. 17:865-872(2000).				
RN	[2]				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C.B17;				
RA	Egan S.E., Cohen B.L., Sarkar M., Ying Y., Cohen S., Singh N.,				
RA	Wang W., Flock G., Goh T., Schachter H.,				
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Breast tumor;				
RA	Strausberg R.,				
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AY043479; AAK95359.1; "				
DR	EMBL; BC009075; AAH09075.1; "				
DR	MGI; MGI:1889505; B3gn1.				
DR	InterPro; IPR002659; Glyco_trans_31.				

```

DR Pfam: PF01762; Galactosyl_T; 1.
KW Glycosyltransferase; transferase.
SQ SEQUENCE 397 AA: 45883 MW: D8BBEA1866C1D106 CRC64;

Query Match 88.2%; Score 1872; DB 11; Length 397;
Best Local Similarity 87.4%; Pred. No. 2.2e-154;
Matches 347; Conservative 19; Mismatches 31; Indels 0; Gaps 0;

Qy 1 MSVGRRRRIKLLGILMANVFIYFIMEVSKSSQKNGKGEVIIPKEFKWKISTPPEAYWN 60
Dy 1 MSVGRRRRVKLLGILMANVFIYLVIVEVSKSSQKNGKGGVILPKEFKWKPPSTPRAYN 60
Qy 61 REQKLNRYNPILSLMTNQTGEAGRLSNISHLNACEPDLRVTSVVTGNNLPDRFKDFL 120
Dy 61 REQKLNRYNPILSLMTNQTGEAGRLSNISHLNACEPDLRVTSVVTGNNLPDRFKDFL 120
Qy 121 LYLRCRNYSLIDOPDKCAKPPFLLAIKSLTPHFARRQAIRESWGQESNAGNQTVVYRVF 180
Dy 121 LYLRCRNYSLIDOPDKCAKPPFLLAIKSLTPHFARRQAIRESWGQESNAGNQTVVYRVF 180
Qy 181 LLGOTPPEDNHPDLSMDLKFESEKHQDILMWNRYDTFFNLSKEVLFRLRWVSTSCPDTEF 240
Dy 181 LLGOTPPEDNHPDLSMDLKFESEKHQDILMWNRYDTFFNLSKEVLFRLRWVSTSCPDTEF 240
Qy 241 VFKGDDDDVFVNTTHILNLSKRAKDLFIGDVIHNAQPHRDKKLYIPEVYSGLY 300
Dy 241 VFKGDDDDVFVNTTHILNLSKRAKDLFIGDVIHNAQPHRDKKLYIPEVYSGLY 300
Qy 301 PPYAGGGGFLYSGHLALRLYHITDQVHLYPIDDDVYTGMCLOKLGVLPEKHKGFRTFDEE 360
Dy 301 PPYAGGGGFLYSGHLALRLYHITDQVHLYPIDDDVYTGMCLOKLGVLPEKHKGFRTFDEE 360
Qy 361 KNKNNICSYVDLMLVHSRKPQEMIDWSQLOSAHLKC 397
Dy 361 KNKNNICSYVDLMLVHSRKPQEMIDWSQLOSPNLKC 397

RESULT 2
Q92222 ID O92222 PRELIMINARY; PRT; 397 AA.
AC O92222;
DT 01-MAR-1999 (TrEMBLrel. 10, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase
DE (EC 2.4.1.149). (Poly-N-acetylglucosamine extension enzyme) (N-
DE acetylglucosaminyltransferase) (Beta3GNT).
GN B3GNT1 OR BETA3GNT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=ICR; TISSUE=Neonatal brain;
RX MEDLINE=99110903; PubMed=9892646;
RA Zhou D., Dinter A., Gutierrez Gallego R., Kamerling J.P.,
RA Villegenthart J.F.G., Berger E.G., Hennet T.;
RT "A beta-1,3-N-acetylglucosaminyltransferase with poly-N-
RT acetylglucosamine synthase activity is structurally related to beta-
RT 1,3-galactosyltransferase."
RL Proc. Natl. Acad. Sci. U.S.A. 96:406-411(1999).
RN [2]
RP REVISIONS.
RA Zhou D., Berger E.G., Hennet T.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CAN INITIATE THE SYNTHESIS OR THE ELONGATION OF THE
CC LINEAR POLY-N-ACETYLGLUCOSAMINOGLYCANS.
CC -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-
CC GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-
CC GLUCOSAMINYL-1,3-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.
CC -!- COFACTOR: MANGANESE.
CC -!- PATHWAY: GLYCOSYLATION.

```

CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, LUNG, KIDNEY AND
CC TESTIS AND TO A LESSER EXTENT IN LIVER AND SKELETAL MUSCLE. NO
CC EXPRESSION IN SPLEEN.
DR EMBL: AF092050; AAD09763.2; -;
DR MGD: MGI:1889505; B3gnl1.
DR InterPro: IPR002659; Galactosyl_T; 1.
DR Pfam: PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW Glycoprotein; Golgi stack.
FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 29 325 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 397 AA: 45813 MW: 29BBEADF7B770A6B CRC64;
Query Match 87.9%; Score 1866; DB 11; Length 397;
Best Local Similarity 87.2%; Pred. No. 7.2e-154;
Matches 346; Conservative 19; Mismatches 32; Indels 0; Gaps 0;
Qy 1 MSVGRRRRIKLLGILMANVFIYFIMEVSKSSQKNGKGEVIIPKEFKWKISTPPEAYWN 60
Dy 1 MSVGRRRRVKLLGILMANVFIYLVIVEVSKSSQKNGKGGVILPKEFKWKPPSTPRAYN 60
Qy 61 REQKLNRYNPILSLMTNQTGEAGRLSNISHLNACEPDLRVTSVVTGNNLPDRFKDFL 120
Dy 61 REQKLNRYNPILSLMTNQTGEAGRLSNISHLNACEPDLRVTSVVTGNNLPDRFKDFL 120
Qy 121 LYLRCRNYSLIDOPDKCAKPPFLLAIKSLTPHFARRQAIRESWGQESNAGNQTVVYRVF 180
Dy 121 LYLRCRNYSLIDOPDKCAKPPFLLAIKSLTPHFARRQAIRESWGQESNAGNQTVVYRVF 180
Qy 181 LLGOTPPEDNHPDLSMDLKFESEKHQDILMWNRYDTFFNLSKEVLFRLRWVSTSCPDTEF 240
Dy 181 LLGOTPPEDNHPDLSMDLKFESEKHQDILMWNRYDTFFNLSKEVLFRLRWVSTSCPDTEF 240
Qy 241 VFKGDDDDVFVNTTHILNLSKRAKDLFIGDVIHNAQPHRDKKLYIPEVYSGLY 300
Dy 241 VFKGDDDDVFVNTTHILNLSKRAKDLFIGDVIHNAQPHRDKKLYIPEVYSGLY 300
Qy 301 PPYAGGGGFLYSGHLALRLYHITDQVHLYPIDDDVYTGMCLOKLGVLPEKHKGFRTFDEE 360
Dy 301 PPYAGGGGFLYSGHLALRLYHITDQVHLYPIDDDVYTGMCLOKLGVLPEKHKGFRTFDEE 360
Qy 361 KNKNNICSYVDLMLVHSRKPQEMIDWSQLOSAHLKC 397
Dy 361 KNKNNICSYVDLMLVHSRKPQEMIDWSQLOSPNLKC 397

RESULT 3
Q92222 ID Q8UWMO PRELIMINARY; PRT; 406 AA.
AC Q8UWMO;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Beta-3-galactosyltransferase.
GN SSP5.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Chou C.-M., Leu J.-H., Huang C.-J.;
RT "A family of novel genes encoding beta-3-galactosyltransferase from
RT zebrafish, zssp5 mRNA.";

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF321831; AAL32299.1; -

DR InterPro; IPR002659; Glyco_trans_31.

DR Pfam; PF01762; Galactosyl_T; 1.

DR Trasnferase; Glycosyltransferase.

SW SEQUENCE 406 AA: 46953 MW: F928471BBDF0F1681 CRC64;

Query Match 58.1%; Score 1232.5; DB 13; Length 406;

Best Local Similarity 58.2%; Pred. No. 8.9e-99;

Matches 226; Conservative 70; Mismatches 87; Indels 5; Gaps 3;

QY 5 RRRKLLGILMANNVFIYEMYSKSSOENKNGEV:IPKEFKWIKTSTPEAYWREQE 64

DB 5 RRVKVVVA--WMTWVFLFIVVSRNAGKSSKNSLPLKRFPAWKDLPDSDAYNRQQQ 62

QY 65 KLNROYNPILSMITNOTGE--AGRLSNISHINCEPDRLRVTSVVTGNNLPDRFKDFLL 122

DB 63 QINYINNRLEKL-NYTDNLDPWLDNTVSLDSCDPDYRTTQVKDYNSLPDRFKDFLL 121

QY 123 LRCRNYSLIIDOPDKCAKPFLLAIKSLTPHFARROAIRSNGQESNAGNQTVVRVFL 182

DB 122 MRCRSYPIVDVDPQNCIKKQPFLLAIKSLPHEDRRQAIRSNGKVGRIANRSVTVFLL 181

QY 183 GQTPEDNHPDLSMLAFSEKSHODILMNNYRTFFNLSKEVLFRLRWVSTSCPDERVF 242

DB 182 GNAATEDHFDPLSKMLHSESHRDILOWDYRTDFNLTKEVLFLEWLSRCPGANFIF 241

QY 243 KGDDOVFNTHILNLSKTKAKDLFGDVITHNAGPHRDKLKYIPEVYSGLYPP 302

DB 242 KGDDOVFNTHIITDFTNLNNAKARELFVGDVITNAGPHRDKKVKYIFESFVGMYP 301

QY 303 YAGGGGLYSCHLALRYHITDOVHLYPIDVYTGMCLOKGLVPEKHKGFRFTDEEKN 362

DB 302 YAGGGGLYSCHLALRYHITDOVHLYPIDVYTGMCLOKGLVPEKHKGFRFTDEEKN 361

QY 363 KNNICSYVDMLVHRSRKPQEMIDWISQL 390

DB 362 RDNACAYKSLMLVHRSRKPQEMIKIWAWL 389

RESULT 4

Q8UWM3

ID Q8UWM3 PRELIMINARY; PRT: 412 AA.

AC Q8UWM3;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Beta-3-galactosyltransferase.

GN SSP2.

OS Brachydanio rerio (zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OX NCBI_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RA Chou C.-M., Leu J.-H., Huang C.-J.;

RT "A family of novel genes encoding beta-3-galactosyltransferase from

zebrafish, zssp3 mRNA."

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF321828; AAL32296.1; -

DR InterPro; IPR002659; Glyco_trans_31.

DR Pfam; PF01762; Galactosyl_T; 1.

DR Trasnferase; Glycosyltransferase.

SW SEQUENCE 412 AA: 47440 MW: 14F24C26570B5F45 CRC64;

Query Match

Best Local Similarity 54.0%; Score 1145.5; DB 13; Length 412;

Matches 224; Conservative 58; Mismatches 111; Indels 7; Gaps 5;

QY 1 MSVGRRRRLKLGILMANNVFIYEMYSKSSQEKNG-KGEVTPREKFKHKTSTPEAYW 59

DB 1 MSNNWNSTKLGVLMLANFLIYIAVEVSRSHNDRNAFKNNKIPKE-FWKTKVSTAEP 59

QY 60 NREOEKLRQYNPILSMITNOTGEAGRLSNISH-LNY-CEPDLRVTSVVTGNNLPDRFK 117

DB 60 NREOEKLRDDIY---YLPVNASSELPFRHGPRLNLHSHCKHDVNVTAQIEDFNSLPORFQ 116

QY 118 DFLLYLRCRNYSLIIDOPDKCAKPFLLAIKSLTPHFARROAIRSNGQESNAGNQTVV 177

DB 117 DFLLYLRCRNYSLIIDOPDKCAKPFLLAIKSLTPHFARROAIRSNGQESNAGNQTVV 176

QY 178 RVFLLGQTPEDNHPDLSMLAFSEKSHODILMNNYRTFFNLSKEVLFRLRWVSTSCP 237

DB 177 TVFLLGNTASTDHFDPDLSNWKHEAELVGDVLOWDYRTDFNLTKEVLFLEWLSRCP 236

QY 238 TEFEVKGDDVFNTHILNLSKTKAKDLFGDVITHNAGPHRDKLKYIPEVYSGLYPP 297

DB 237 AQYVEKGGDDVFNTHILNLSKTKAKDLFGDVITHNAGPHRDKLKYIPEVYSGLYPP 296

QY 298 GLYPYAGGGGLYSCHLALRYHITDOVHLYPIDVYTGMCLOKGLVPEKHKGFRFT 357

DB 297 GGYPPYAGGGGLYSCHLALRYHITDOVHLYPIDVYTGMCLOKGLVPEKHKGFRFT 356

QY 358 IEKNNKNCISYVDMLVHRSRKPQEMIDWISQLSAHLK 397

DB 357 IEKNNKNCISYVDMLVHRSRKPQEMIDWISQLSAHLK 396

RESULT 5

Q8UWM2

ID Q8UWM2 PRELIMINARY; PRT: 412 AA.

AC Q8UWM2;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Beta-3-galactosyltransferase.

GN SSP3.

OS Brachydanio rerio (zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OX NCBI_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RA Chou C.-M., Leu J.-H., Huang C.-J.;

RT "A family of novel genes encoding beta-3-galactosyltransferase from

zebrafish, zssp3 mRNA."

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF321829; AAL32297.1; -

DR InterPro; IPR002659; Glyco_trans_31.

DR Pfam; PF01762; Galactosyl_T; 1.

DR Trasnferase; Glycosyltransferase.

SW SEQUENCE 412 AA: 47812 MW: 80C554675A2CB00A CRC64;

Query Match

Best Local Similarity 32.1%; Score 680.5; DB 13; Length 412;

Matches 123; Conservative 66; Mismatches 100; Indels 5; Gaps 2;

QY 109 FNNLPDRKDFLLYLRCRNYSLIIDOPDKCAKPFLLAIKSLTPHFARROAIRSNGQ 168

DB 113 FRLSPRHEFVLRHRCRYFPMNLNHPKCGGVDVLLVVKSVTEEDHRRRAVRKTGKE 172

QY 169 SNAGNQTVVRVFLGQTPEDNHPDLSMLAFSEKSHODILMNNYRTFFNLSKEVLF 228

DB 173 QEIOGLKIKTFLFLLGTPAPGKDSRLQALQVYEDRTYGDILQWDFMDTFFNLTKEVNF 232

QY 229 RWVSTSCPDETFVFKGDDVFNTHILNLSKTKAKDLFGDVITHNAGPHRDKL 287

DB 233 RWFSTYCPDVPFIFKGGDDVFNTHILNLSKTKAKDLFGDVITHNAGPHRDKL 292

QY 288 KYIYIPEVYSGLYPPYAGGGGLYSCHLALRYHITDOVHLYPIDVYTGMCLOKGLV 347

DB 293 KYIYIPEVYSGLYPPYAGGGGLYSCHLALRYHITDOVHLYPIDVYTGMCLOKGLV 352

QY 348 EXHKGFRFTDIEEKNKNNI---CSYVDMLVHRSRKPQEMIDWISQLSAHLK 397


```
D6 109 FQSLPFRKFLAYRCHRCVFPMLNHPKCAQGVMLVVKSVITQDREVRIRQTWGH 156
QY 169 ---SNAGNQTVVRVFLGQTPTPDNDHPDLSDMLKFESEKHQDILMNYKDTFFNLKKEV 225
DB 157 WESAGLGRGAVRFLFLGTASKQERTHYOQLLAYEDRLYADILQWDFLDSFNLTKEL 216
QY 226 LFLRWSTSPOTFEVFKGDDDFVFNTHILNLSKTKAKDPLFGDVIHNAHPHROK 285
DB 217 HFLKWLDIYCPNVPFVFKGDDDFVFNTHILNLSKTKAKDPLFGDVIHNAHPHROK 274
QY 286 KLXYIPEVYVY-SGLYPPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYGMCKLQKLG 344
DB 275 DNYIYIPAVNYGATVPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYGMCKLQKLG 334
QY 345 LVPEKHKGFTFTEIEKNKNNI----CSYVDLMLVHSRKPQEMIDVWSQLOSAHLKC 397
DB 335 VKPTGHEGFTFGISRVRSSRMNKEPCFYRAMLVVHKLLPAELLAMWDLVHS-NLTC 390

RESULT 9
Q8UWM4 PRELIMINARY: PRT: 418 AA.
AC Q8UWM4
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Beta-3-galactosyltransferase.
GN SSPI.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Chou C.-M., Liu J.-H., Huang C.-J.;
RT "A family of novel genes encoding beta-3-galactosyltransferase from
RT zebrafish, zsspl mRNA";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF321827; AAL32295.1; -.
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 418 AA; 48936 MW; C7225B6259E1622B CRC64;

Query Match 30.78; Score 651; DB 13; Length 418;
Best Local Similarity 43.18; Pred. No. 3.3e-48;
Matches 127; Conservative 61; Mismatches 97; Indels 10; Gaps 5;

QY 109 FNNLPDRFKDFLLYLCRCNYSLLIDOPDKCAKPFLLLAIAKSLTPHARRQAIRESWGQE 168
DB 120 FTGPDNFKOFLYRHCYRFPMLNHPKCSGEIDLLIVIKSVITQDREVRIRKTWGRE 179
QY 169 SNAGNQTVVRVFLGQTPTPDNDHPDLSDMLKFESEKHQDILMNYKDTFFNLKKEVIF 228
DB 180 QVLNGKRIKTLFLLGKSSNLEERANHQKLEYEDYIYGDVLQWDFMDSFNLTKELHFL 239
QY 229 RNVSTSPOTFEVFKGDDDFVFNTHILNLSKTKAKDPLFGDVIHNAHPHROK 288
DB 240 KWFSSYCPKQIYIPKGGDDDFVSNPIFEYLE--ISGNLKDPLFVGDFVLFKAKPIREQNK 297
QY 289 YIPEVYVY-SGLYPPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYGMCKLQKLG 347
DB 298 YIIPQALYNKTLIPYAGGGFLMDGALKARKLYGACETLELPIDDVFLGMCLEVLQVTP 357
QY 348 EKHKGFTFTEIEKNKNNI----ICSYVDLMLVHSRKPQEMIDVWSQLOSAHLKC 397
DB 358 IKHNAFTFGL-VKNKTSRLNRPFCFFKSLIVVHKLLPPDPLMSMW-KLVNSDLIC 410
```

RESULT 10
Q923H4

```
ID Q923H4 PRELIMINARY: PRT: 350 AA.
AC Q923H4
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Bctal;3 N-acetylglucosaminyltransferase-4.
GN B3GNT4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhou D., Hennes T.;
RT "Mouse Bctal;3 N-acetylglucosaminyltransferase-4.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY037786; AAK68856.1; -.
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 350 AA; 39887 MW; 55A12083026D29B1 CRC64;

Query Match 29.9%; Score 635.5; DB 11; Length 350;
Best Local Similarity 41.5%; Pred. No. 5.9e-47;
Matches 130; Conservative 59; Mismatches 105; Indels 19; Gaps 8;

QY 91 SHLNYCEPDLRVTSVYTFNNLPDRFKDFLLYLCRCNYSLLIDOPDKCAKPFLLLAIAKS 150
DB 45 SHHSQSPNL--TVNNTSLPSLRHRLFLTYRCHNFSILL-EPSECAKDTFLLVKIS 100
QY 151 LTPHARRQAIRESWGQESN-AGNQTVYRVFLL---GQTPEDNHPDLSDMLKFESEKHQ 206
DB 101 QPAHIEQSAIRSTWGRAGSARGQLKLVFLLVAGVPVP-----AQLLVYESWQDF 153
QY 207 DILMNYRDTFFNLKKEVLEFLRWYVSTSCPDTEFVKGDDDFVFNTHILNLSKSKTK 265
DB 154 DILQWDFAEDEFNLTKELHVRQWIAAATACTAHFLKGGDDDFVHVPNVLFFLEGWD--P 211
QY 267 AKDLFGDVIHNAHPHROKLLKYIPEVYVYSGL-YPPYAGGGFLYSGHLALRLYHITDQ 325
DB 212 AQDFLVGDVIRLARENRTKVFYFPMYRARIHYPPYAGGGGYVMSQATVYRHLHRAEMEE 271
QY 326 VHLXPIDDVYGMCKLQKLVPERKHKGRTEDIEK-NKNNICSYVDLMLVHSRKPQEMI 384
DB 272 AELFPIDDVYGMCKLRLKLVTPIHAGFKTFGIQOPLNRPDPCLYKGLLVHRLSPLEMW 331
QY 385 DIWSQLOSAHLKC 397
DB 332 TWWALVTDERLKC 344

RESULT 11
Q8N934 PRELIMINARY: PRT: 353 AA.
AC Q8N934
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein FLJ38427.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissue=Brain;
RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N.,
RA Matsumoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,
RA Matsumura Y., Moriya S., Chiba E., Momiya H., Onogawa S.,
RA Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Kataoka R.,
RA Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terashima Y.,
RA Watanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A.,
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:54:34 ; Search time 12.5 seconds
(without alignments)
1493.568 Million cell updates/sec

Title: US-09-804-357B-2

Perfect score: 2129

Sequence: 1 MSVGRRRYKLLGILMHANVF.....RKQEMIDWSQLSPNLKC 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1862	87.5	397	1 B3G7_HUMAN	Q9ny97 h beta-1,3-
2	619.5	29.1	372	1 B3G8_HUMAN	Q9y2a9 h beta-1,3-
3	400	18.8	308	1 B3G5_MOUSE	Q9j167 m beta-1,3-
4	354	16.6	300	1 B3G5_GORGO	Q9s293 g beta-1,3-
5	354	16.6	301	1 B3G5_PANPA	Q9s294 p beta-1,3-
6	353	16.6	310	1 B3G5_HUMAN	Q9y263 h beta-1,3-
7	349	16.4	297	1 B3G5_PANTR	Q9s295 p beta-1,3-
8	314.5	14.8	371	1 B3G4_MOUSE	Q9z0f0 mus musculus
9	310.5	14.6	371	1 B3G4_RAT	O88178 rattus nerv
10	287	13.5	378	1 B3G4_HUMAN	O96024 homo sapien
11	104.5	4.9	1026	1 BGAL_STRTR	P23989 streptococc
12	103.5	4.9	610	1 RCK2_YEAST	P38623 saccharomyc
13	100	4.7	432	1 BPER_THEMA	O33833 thermotoga
14	99	4.7	410	1 FENR_ANASO	P21890 anabaena sp
15	99	4.7	517	1 VLI_HPV25	Q02051 human papil
16	98	4.6	1018	1 YC14_METJA	Q58611 methanococc
17	96	4.5	2733	1 RRPB_CYMA5	P16342 murine coro
18	95	4.5	626	1 HTPG_BUCBP	O89a93 buchnera ap
19	94.5	4.4	800	1 PI27_YEAST	P32606 saccharomyc
20	94	4.4	472	1 KSGT_ARATH	Q96287 arabidopsis
21	93	4.4	440	1 FENR_ANAVA	Q4549 anabaena va
22	92.5	4.3	495	1 SYE2_THETN	O887t1 thermocaul
23	92.5	4.3	995	1 SR13_MOUSE	Q923q2 mus muscu
24	92	4.3	440	1 FENR_ANASP	P58558 anabaena sp
25	89.5	4.2	842	1 DP3A_THEMA	Q92hg4 thermotoga
26	89.5	4.2	2131	1 YCF2_SPIOL	P08973 spinacia ol
27	89	4.2	468	1 KSGT_BRANA	O04160 brassica na
28	89	4.2	597	1 TTPI_YEAST	P38069 saccharomyc
29	88.5	4.2	249	1 RL7A_SCHPO	P17937 schizosacch
30	88.5	4.2	4466	1 DYHC_TRIGR	P23098 tripneustes
31	87.5	4.1	498	1 GALT_STACA	O9tgr9 staphylococ
32	87.5	4.1	1024	1 Y075_MYCGE	P47321 mycoplasma
33	87	4.1	705	1 PPK_BACHD	Q9kd27 bacillus ha

RESULT 1
B3G7_HUMAN
ID B3G7_HUMAN STANDARD; PRT: 397 AA.
AC Q9NY97; Q9NQ99; Q9NQ90; Q9NUT9;
DI 16-OCT-2001 (Rel. 40, Created)
DI 15-SEP-2003 (Rel. 42, Last sequence update)
DI 15-SEP-2003 (Rel. 42, Last annotation update)
DE Beta-1,3-galactosyltransferase 7 [EC 2.4.1.-] (Beta-1,3-GalTase 7)
DE (Beta3Gal-T7) (b3Gal-T7) (UDP-galactose:beta-N-acetylglucosamine beta-1,3-galactosyltransferase 7) (UDP-Gal:beta-GlcNAc beta-1,3-galactosyltransferase 7) (Beta-3-Gx-T7).
GN B3GNT1 OR B3GALT7.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Amado M., Carneiro F., Clausen H.;
RT "Cloning and expression of two beta-1,3-galactosyltransferases:
beta3gal-T5 and beta3gal-T6.";
RI Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Urinary bladder;
KA Gromova I., Gromov P., Celis J.E.;
RI "A novel member of beta-1,3-galactosyltransferase family is down
regulated during bladder TCC progression.";
RI Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain, and Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
KA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Hopkins R.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ussid T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.C., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RI Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 25-397 FROM N.A.
RC TISSUE=Placenta;

34 67 4.1 838 1 SYV_MYCPN P75304 mycoplasma
35 87 4.1 1427 1 MES4_DROME Q8mt36 dtosophilla
36 87 4.1 4036 1 RRPL_DUGBV Q66431 dugbe virus
37 86.5 4.1 995 1 SMT3_HUMAN P87207 candida alb
38 86 4.0 378 1 YKDA_MYCCA P45615 mycoplasma
39 85.5 4.0 655 1 YCB9_YEAST P25384 saccharomyc
40 85.5 4.0 1770 1 VSX2_BRARE O42477 brachydanio
41 85 4.0 395 1 APEA_THEMA O9wyj9 thermotoga
42 84.5 4.0 451 1 CPTM_RAT Q63704 rattus norv
43 84.5 4.0 772 1 RIR1_NEOCR Q9uw15 neurospora
44 84.5 4.0 929 1 SYI_AQUAE O66651 aquifex aeo
45 84.5 4.0 956 1

QY	121	LYLCRNVSLLDIPDKCAKCAKPELLLAIKSLIPHFARQAIRESGRTNVTGNQTVVRVF	180
DB	121	LYLCRNVSLLDIPDKCAKCAKPELLLAIKSLIPHFARQAIRESGQESNAGNQTIVRVF	180
QY	181	LLGKTPPDNHPDLSMDLKAFESDKHODILMWNVROTFFNLSLKEVLFURWYSTSCPDAEF	240
DB	181	LLGQTPPDNHPDLSMDLKAFESDKHODILMWNVROTFFNLSLKEVLFURWYSTSCPOTEF	240
QY	241	VFKGDVDFVNTTHILNLSLSKSKAKDLFGDVIHNAHPHROKRLKYYIPEVFIYGVY	300
DB	241	VFKGDVDFVNTTHILNLSLSKSKAKDLFGDVIHNAHPHROKRLKYYIPEVVISGLY	300
QY	301	PPYAGGGFLYSGPALLRLYSATSRVRLYPIDDVYTGMCLOKLGIVPEKHKGFRTDIEE	360
DB	301	PPYAGGGFLYSGHGLAURLYHTIQVHLVPIDDVYTGMCLOKLGIVPEKHKGFRTDIEE	360
QY	361	KNKNKNCISYDMLVHRSKPKQEMIDWSQLQSPNLKC	397
DB	361	KNKNKNCISYDMLVHRSKPKQEMIDWSQLQSAHLKC	397
RESULT 2			
ID	B3G8_HUMAN	STANDARD:	PRT; 372 AA.
AC	C9Y2A9;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Beta-1,3-galactosyltransferase 8 (EC 2.4.1.-) (Beta-1,3-GalTase 8)		
DE	(beta3Gal-T8) (b3Gal-T8) (UDP-galactose:beta-N-acetylglucosamine beta-		
DE	1,3-galactosyltransferase 8) (UDP-Gal:beta-GlcNAc beta-1,3-		
DE	galactosyltransferase 8) (Beta-3-Gx-T8) (Core 1 extending beta-1,3-N-		
DE	acetylglucosaminyltransferase) (Core1-beta3GlcNAcT).		
GN	B3GNT3 OR B3GALT8 OR TWEM3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=99173880; PubMed=10072769;		
RX	Yokoyama-Kobayashi M., Yamaguchi T., Sekine S., Kato S.;		
RA	"Selection of cDNAs encoding putative type II membrane proteins on the		
RA	cell surface from a human full-length cDNA bank.;"		
RL	Cell 228:161-167(1999).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=21332592; PubMed=11439191;		
RX	Yen J., Hiraoka N., Petryniak B., Nakayama J., Ellies L.G., Rabuka D.,		
RA	Hardgaul O., Marth J.D., Lowe J.B., Fukuda M.;		
RA	"Novel sulfated lymphocyte homing receptors and their control by a		
RL	core1 extension beta1,3-N-acetylglucosaminyltransferase.;"		
RL	Cell 105:957-969(2001).		
RL	[3]		
RP	SEQUENCE FROM N.A.		
RP	Jensen M.A., Bennett E.P.;		
RA	"Cloning of a new member of the beta 1,3 Galactosyltransferase family,		
RL	b1,3Gal-T6.;"		
RL	Submitted (NOV-1998) to the ENBL/GenBank/DBJ databases.		
RL	[4]		
RP	REVIEW.		
RP	MEDLINE=20047730; PubMed=10580128;		
RX	Anado M., Almeida R., Schwientek T., Clausen H.;		
RA	"Identification and characterization of large galactosyltransferase		
RA	gene families: galactosyltransferases for all functions.;"		
RL	Biochim. Biophys. Acta 1473:35-53(1999).		
CC	-1- SUBWAY: Glycosylation.		
CC	-1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).		
CC	-1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

ENBL: AB015630; BAA76497.1; -
 DR ENBL: AF293373; AAK00849.1; -
 DR ENBL: AJ130847; CAC45044.1; -
 DR GEN: HGNC:13528; B3GNT3.
 DR GO: GO:0005887; C:integral to plasma membrane; TAS.
 DR InterPro: IPR002659; Glyco_Tran_31.
 DR Pfam: PF01762; Galactosyl_T_1.
 DR Transferrase, Glycosyltransferase; Glycoprotein; Transmembrane;
 KW Signal-anchor; Golgi stack; Multigene family.
 FT DOMAIN 1 10 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 11 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT FT DOMAIN 32 372 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 372 AA; 42515 MW; A63FE88AFCFC62 CRC64;

Query Match 29.1%; Score 619.5; DB 1; Length 372;
 Best Local Similarity 38.1%; Pred. No. 6.3e-44;
 Matches 135; Conservative 53; Mismatches 133; Indels 33; Gaps 7;

QY 50 KPSTPRAYNRSQEKLNRWNPILNRVANQGTGELATSPNTSHLSYCEPDSTVMTAVTDF 109
 :||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 39 QPPAIPAL-----AMPTP-----PTRPAPAPCHAMTSMVTH--PDF 73

QY 110 NNLPDRFKDLYLCRNYSLLID-OPKKCAKPKPFLLLAJKSLIPHFARQALRESMGRE 168
 :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 74 ATQPHQVONFLLYRHCHEPFLLODVPSPCAQVPYFLLLVKSSPSNVVRELLRRTWGRE 133

QY 169 TNYGNQTVRVFLVLTGTPPDNHPDLSDMUKFSDKHQDILMNNYRDTFFNLSLKEVFLF 228
 : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 134 RKYRGLQLRLFLVGTASNPHEARKYNRLLEALEAQHTGDILOWDFHDSFFNLTLKQVFLF 193

QY 229 RMYSTSCPDAEEFYKGDGDDVFVNTHTILNLSLSKSKAKDLFTGDIHAGPHROKKLK 268
 :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 194 QWQETRCANASFVLNGDDDDVFATDNWVFLQ--DHDGPHLFGVQLIQNVGPIRAFWSK 251

QY 289 YIPEVFTYGV-YPPYAGGGGFLYSPALLRLYSATSRKVLHYPIDQVYTCMLQKLGCVF 347
 :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 252 YYPEVFTYGVYPPYGGGGFLSRFTAAALRRAAHVLDIFFDDVFLGCMCLEGLKLP 311

QY 348 EKHKGRFTFDIEKNKK---NTCSYIDLMLVHSRKPQEMIDTWSQLSPNLKC 397
 :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 312 ASHSGIRTSQVRAPSDQHLSSFDPCFYRDLLLVHRLDPYFMLLMWALNQPNLTC 365

RESULT 3
 B3G5_MOUSE ID B3G5_MOUSE STANDARD: PRT: 308 AA.
 AC QJ167; 2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
 DE (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
 DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
 DE galactosyltransferase 5) (Beta-3-Gx-T5) (Stage-specific embryonic
 DE antigen-3 synthase) (SSEA-3 synthase).
 GN B3GALT5 OR B3GT5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.


```

DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.1-) (Beta-1,3-GalTase 5)
DE (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
GN B3GALT5.
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX NCBI_TaxID=9595;
RN [1]
RP SEQUENCE FROM N.A.
RL Liu Y., Saitou N.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOPOLIPID LC3CER AS AN
CC EFFICIENT ACCEPTOR (BY SIMILARITY).
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB041416; BAA94501.1;
DR InterPro: IPR002659; Glyco_trans_31.
DR Pfam: PF01762; Galactosyl_T. 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack; Multigene family.
FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 29 >300 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 130 130 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 300 AA; 35104 MW; 1AA11692ED9F06FA CRC64;

Query Match 16.6%; Score 354; DB 1; Length 300;
Best Local Similarity 30.3%; Pred. No. 4.7e-22;
Matches 76; Conservative 56; Mismatches 111; Indels 8; Gaps 4;

Qy 142 PFLLLAIAKSLIPHARRQAIRESWGRETNGQTVVRFLLGKTPPDNHPDLSMLKFE 201
Db 57 PFLVLLVTSKQLAERMAIROTGWKERTYKQKLTFFLLGTTSSAAETREVDQ---E 112
Qy 202 SDKHQDILMNNYRDTFFNLSLKEVLFLRWVSTSCPDAEFVFGDDVFNTHILNLYNS 261
Db 113 SQRHGDIIOKDFLDVYNTLTMTMGIEWVHRFCQAAAFVMTKDSDFINVDYLTLL-- 170
Qy 262 LSKXAKDPIGVDIHNAHGRDKKLYYIPEVFT-GVYPYAGGGFLYSGPALLRLY 320
Db 171 LKKNRTTRFTGFLKLNFEPIRQPSKFWFSKSEYFWDYPPFCSGTYGVSGDVASQVY 230
Qy 321 SATSRVHLYPIDDDVYTGMLQKLGVLPEKHGKGFRTFDEEKNKNKICSYIDLMLVHSRKP 380
Db 231 NVSKSPYIKLEDVFGVGLCLERLNIRLELHQSPTF-FPGGLRFSVCVLRFRIVACHFIKP 289
Qy 381 QEMIDIWSQIQ 391
Db 290 RTLLDYWAOLE 300

RESULT 5
B3G5_PANPA
ID B3G5_PANPA PRT; 301 AA.
AC Q9N294;

```

```

DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.1-) (Beta-1,3-GalTase 5)
DE (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
GN B3GALT5.
OS Pan paniscus (Pygmy chimpanzee) (Bonobo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9597;
RN [1]
RP SEQUENCE FROM N.A.
RL Liu Y., Saitou N.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOPOLIPID LC3CER AS AN
CC EFFICIENT ACCEPTOR (BY SIMILARITY).
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB041415; BAA94500.1;
DR InterPro: IPR002659; Glyco_trans_31.
DR Pfam: PF01762; Galactosyl_T. 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack; Multigene family.
FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 29 >301 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 130 130 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 301 AA; 35222 MW; E04CC70F41CB29FD CRC64;

Query Match 16.6%; Score 354; DB 1; Length 301;
Best Local Similarity 30.2%; Pred. No. 4.8e-22;
Matches 76; Conservative 57; Mismatches 111; Indels 8; Gaps 4;

Qy 142 PFLLLAIAKSLIPHARRQAIRESWGRETNGQTVVRFLLGKTPPDNHPDLSMLKFE 201
Db 57 PFLVLLVTSKQLAERMAIROTGWKERTYKQKLTFFLLGTTSSAAETREVDQ---E 112
Qy 202 SDKHQDILMNNYRDTFFNLSLKEVLFLRWVSTSCPDAEFVFGDDVFNTHILNLYNS 261
Db 113 SQRHGDIIOKDFLDVYNTLTMTMGIEWVHRFCQAAAFVMTKDSDFINVDYLTLL-- 170
Qy 262 LSKXAKDPIGVDIHNAHGRDKKLYYIPEVFT-GVYPYAGGGFLYSGPALLRLY 320
Db 171 LKKNRTTRFTGFLKLNFEPIRQPSKFWFSKSEYFWDYPPFCSGTYGVSGDVASQVY 230
Qy 321 SATSRVHLYPIDDDVYTGMLQKLGVLPEKHGKGFRTFDEEKNKNKICSYIDLMLVHSRKP 380
Db 231 NVSKSPYIKLEDVFGVGLCLERLNIRLELHQSPTF-FPGGLRFSVCVLRFRIVACHFIKP 289
Qy 381 QEMIDIWSQIQS 392
Db 290 RTLLDYWAOLEN 301

RESULT 6

```



```

RESULT 8
B3G4_MOUSE STANDARD; PRT; 371 AA.
AC Q920F0; Q91VC1; Q920U8; Q920U9;
CT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Beta-1,3-galactosyltransferase 4 (EC 2.4.1.62) (Beta-1,3-GalTase 4)
DE (Beta3galT4) (B3Gal-T4) (Ganglioside galactosyltransferase) (UDP-
DE galactose-beta-N-acetyl-galactosamine-beta-1,3-galactosyltransferase)
DE (GAL-T2).
DE B3GALT4.
DE GN
DE OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1] SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL
RP STAGE.
RC MEDLINE-99432136; PubMed-10502288;
RA Danlotti J.L., Martina J.A., Zurita A.R., Maccioni H.J.F.;
RA "Mouse beta 1,3-galactosyltransferase (Gal/GM1/GD1b synthase): protein
RT characterization, tissue expression, and developmental regulation in
RT neural retina.";
RL J. Neurosci. Res. 58:318-327(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-129/SVJ;
RA Rowen L., Qin S., Madan A., Loretz C., Hall J., James R., Dors M.,
RA Shaffer T., Abbasi N., Ratcliffe A., Dickhoff R., Lasky S., Hood L.;
RA "Sequence of the mouse major histocompatibility complex class II
RT region.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-370 FROM N.A.
RC STRAIN-BLUG2/Msf, BFM/2Msf, C57BL/10SnJ, CAST/Ei, HMI/Msf, MSM/Msf,
RC NUL/Msf, Pgn2, and SWN/Msf;
RA Liu Y., Kitano T., Koide T., Shiroishi T., Moriaki K., Saitou N.;
RA "Conspicuous differences among gene genealogies of 21 nuclear genes of
RT five Mus musculus subspecies.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Involved in GM1/GD1b/GAL ganglioside biosynthesis.
CC -!- CATALYTIC ACTIVITY: UDP-galactose + N-acetyl-D-galactosaminyl-(N-
CC + D-galactosyl)-D-galactosyl-D-glucosyl-N-acetylneuraminyl)-D-
CC galactosyl-D-glucosyl-N-acetylneuraminyl)-D-
CC galactosyl-D-glucosyl-N-acetylneuraminyl)-D-
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
CC -!- TISSUE SPECIFICITY: Expressed in heart, brain, spleen, kidney,
CC lung and testis.
CC -!- DEVELOPMENTAL STAGE: First expressed at embryonic day 3.
CC Maintained at high levels between days 4 and 7 and declines
CC thereafter to stabilize at low levels after day 10.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AF110520; AAC97977.1; -
DR EMBL: AF082504; AAC69622.1; -
DR EMBL: AF100950; AAC69897.1; -
DR EMBL: AB039164; BAB68688.1; -
DR EMBL: AB039165; BAB68689.1; -
DR EMBL: AB039167; BAB68691.1; -
DR EMBL: AB039168; BAB68692.1; -

```

DR EMBL; AB039170; BAB68694.1; -
 DR EMBL; AB039171; BAB68695.1; -
 DR EMBL; AB039172; BAB68696.1; -
 DR EMBL; AB039173; BAB68697.1; -
 DR EMBL; AB039174; BAB68698.1; -
 DR EMBL; AB039175; BAB68699.1; -
 DR MGD; MGI:1859517; B3gal4.4.
 DR InterPro; IPR002659; Glyco_trans_31.
 DR Pfam; PF01762; Galactosyl_T; 1.
 KW Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane;
 FT Signal-anchor; Golgi stack; Polymorphism.
 FT DOMAIN 1 4
 FT TRANSFER 5 25
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT LUMENAL, CATALYTIC (POTENTIAL).
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARIANT 248 248 W -> R (IN STRAIN PGN2).
 FT VARIANT 344 344 G -> W (IN STRAIN BUC2/MSF).
 SQ SEQUENCE 371 AA: 41235 MW; 323A7FFA56B723B3 CRC64;
 Query Match 14.8%; Score 314.5; DB 1; Length 371;
 Best Local Similarity 29.4%; Pred. No. 1.2e-18;
 Matches 88; Conservative 47; Mismatches 119; Indels 45; Gaps 9;
 QY 130 LLIDQPKK---AKKPFLLLAISLIPFAARRQAIRESWGRETNGVQTVVRFVFLGKTP 186
 Db 56 LLISNSHACGGSGPPFLLILVCTAPEHLNQRAIRASWGAIARCFRVTQLFLGK-P 114
 QY 187 PEDNHPDLSMLKFESDKHQDILMNRYDTFFNLSLKEVLEFLRWVSTSCDAEFVFGDD 246
 Db 115 RHOQADLSS-----ESAARHDILOASFDOSYRNLTSLTSLGNWVNYCPMARYILKTD 170
 QY 247 DVFVNTHTLNLV-----NSLSKSKAKD-----LFIGDVIHNAAGPH 282
 Db 171 DVVNVVPELVSELIQGGPSQWQKGEQETTAIHEHRGQAVPLLYLGRVHVRPT 230
 QY 283 RDKLKYYIPEVFT---GVTPPYAGGGFLYSGPALLRLYSATSRVHLYPIDVYTGMC 339
 Db 231 RTPESRRHVSEELWPNMGPPFPYASGTGYVLSISAVQLILKVASRAPPLPLEDFVGV 290
 QY 340 LQKLGVLPE---KHKGFRTFDEEKNKNKICSYIDMLV-HSRKPOEMIDWISLOSPN 394
 Db 291 ARRGGLATHCVKLAGATHYPLDR-----CCYCKFLTSHKVDPMQKAWKLVRGLN 343
 RESULT 9
 B3G4_RAT
 ID B3G4_RAT
 AC O88178
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Beta-1,3-galactosyltransferase 4 (EC 2.4.1.52) (Beta-1,3-GalTase 4)
 DE (Beta3Gal-T4) (b3Gal-T4) (Ganglioside galactosyltransferase) (UDP-
 DE galactose:beta-N-acetyl-galactosamine-beta-1,3-galactosyltransferase)
 DE (GAL-T2).
 GN B3GAL4.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL
 RP STAGE.
 RC TISSUE=Brain;
 RX MEDLINE=97460055; PubMed=9312075;
 RA Miyazaki H., Fukumoto S., Okada M., Hasegawa T., Furukawa K.,
 RA Furukawa K.;
 RT "Expression cloning of rat cDNA encoding UDP-galactose:GD2 beta1,3-
 RT galactosyltransferase that determines the expression of
 RT GD1b/GM1/GAL1";
 RL J. Biol. Chem. 272:24794-24799(1997).
 CC -|- FUNCTION: Involved in GM1/GD1b/GAL1 ganglioside biosynthesis.
 CC -|- CATALYTIC ACTIVITY: UDP-galactose + N-acetyl-D-galactosaminyl-(N-

acetylneuraminy)-(D-galactosyl-D-glucosyl-N-acylsphingosine - UDP
 + D-galactosyl-N-acetyl-D-galactosaminyl-(N-acetylneuraminy))-D-
 galactosyl-D-glucosyl-N-acylsphingosine.
 CC -|- PATHWAY: Glycosylation.
 CC -|- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
 CC -|- TISSUE SPECIFICITY: Highly expressed in thymus, spleen, kidney and
 CC testis and, to a lesser extent, in brain and liver.
 CC -|- DEVELOPMENTAL STAGE: In the embryonic brain, expression begins at
 CC day 12 and continues until birth. Expression is maintained at low
 CC levels in adult brain.
 CC -|- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB003478; BAA32045.1; -
 DR InterPro; IPR002659; Glyco_trans_31.
 DR Pfam; PF01762; Galactosyl_T; 1.
 KW Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane;
 KW Signal-anchor; Golgi stack.
 FT DOMAIN 1 4
 FT TRANSFER 5 25
 FT SIGNAL-ANCHOR (POTENTIAL)
 FT (POTENTIAL).
 FT LUMENAL, CATALYTIC (POTENTIAL).
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 143 143
 FT SEQUENCE 371 AA: 41254 MW; 327FB76FDACF131 CRC64;
 Query Match 14.6%; Score 310.5; DB 1; Length 371;
 Best Local Similarity 28.4%; Pred. No. 2.5e-18;
 Matches 85; Conservative 52; Mismatches 117; Indels 45; Gaps 9;
 QY 130 LLIDQPKK---AKKPFLLLAISLIPFAARRQAIRESWGRETNGVQTVVRFVFLGKTP 186
 Db 56 LLIPNPQACGGSGPPFLLILVCTAPEHLNQRAIRASWGAIARCFRVTQLFLGK-P 114
 QY 187 PEDNHPDLSMLKFESDKHQDILMNRYDTFFNLSLKEVLEFLRWVSTSCDAEFVFGDD 246
 Db 115 MQQGFADLAS-----ESAAGDVLQASFDOSYRNLTSLTGLNWNVNYCPMARYILKTD 170
 QY 247 DVFVNTHTLNLV-----NSLSKSKAKD-----LFIGDVIHNAAGPH 282
 Db 171 DVVNVVPELVSELIQGGPSQWQKGEQETTAIHEHRGQAVPLLYLGRVHVRPT 230
 QY 283 RDKLKYYIPEVFT---GVTPPYAGGGFLYSGPALLRLYSATSRVHLYPIDVYTGMC 339
 Db 231 RTPESRRHVSEELWPNMGPPFPYASGTGYVLSISAVQLILKVASRAPPLPLEDFVGV 290
 QY 340 LQKLGVLPE---KHKGFRTFDEEKNKNKICSYIDMLV-HSRKPOEMIDWISLOSPN 394
 Db 291 ARRGGLATHCVKLAGATHYPLDR-----CCYCKFLTSHKVDPMQKAWKLVRGLN 343
 RESULT 10
 B3G4_HUMAN
 ID B3G4_HUMAN
 AC O96024
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Beta-1,3-galactosyltransferase 4 (EC 2.4.1.62) (Beta-1,3-GalTase 4)
 DE (Beta3Gal-T4) (b3Gal-T4) (Ganglioside galactosyltransferase) (UDP-
 DE galactose:beta-N-acetyl-galactosamine-beta-1,3-galactosyltransferase)
 DE (GAL-T2) (GAL-T4).
 GN B3GALT4 OR GALT4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;


```
Db 96 -----ETAKS-----EGNGKAIFVKIDSGAKAF--AKPFAEQQL 127
QY 189 ---DNHPDLSMDLKFESDKHODILMMNYRDTFFNLSLKEVFLRWVSISCPDAEFVFKGD 245
Db 128 KKDNKGNKTWQAK--AKHADVPVNLRY-----PNAPIGK-- 161
QY 246 DDVEVNTFHILNLSUSKSKAKDLFGDVIHGNAGPHRDKLKYIPEVYTGTVYPPYAG 305
Db 162 --VISN-----EPLVKEGGIGIVQHIKFDLTGSLNKYI--EGQSIGIIPPGVD 205
QY 306 GGGFLYSGPALLRLYSATSRVHLYPIDDDVYTGMLQKL 343
Db 206 KNG-----KPEKRLRYS:ASTRHGDVDDKTSISLCVRQL 239

RESULT 15
VLL_HPV25
ID VLI_HPV25 STANDARD: PRT: 517 AA.
AC Q02051;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Major capsid protein L1.
GN L1
OS Human papillomavirus type 25.
OC Viruses: dsDNA viruses, no RNA stage: Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10609;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94265501; PubMed=8205838;
RA Delius H. Hofmann B.;
RT "Primer-directed sequencing of human papillomavirus types.";
RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).
RN [2]
RP SEQUENCE OF 315-358 FROM N.A.
RX MEDLINE=92407963; PubMed=1326639;
RA Chan S.Y., Bernard H.U., Ong C.K., Chan S.P., Birgit H., Delius H.;
RT "Phylogenetic analysis of 48 papillomavirus types and 28 subtypes and
variants: a showcase for the molecular evolution of DNA viruses.";
J. Virol. 66:5714-5725(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M96283; AAA47022.1; -
DR EMBL; X74471; CAA52529.1; -
DR EMBL; M96289; AAA47028.1; -
DR PIR; S36496; S36496.
DR InterPro: IPR002210; PV_capsid_L1.
DR Pfam: PF00500; late_protein_L1; 1.
DR PRINTS; PR00865; HPVcapsIDL1.
DR ProDom; PD000544; PV_capsid_L1; 1.
KW Coat protein; Late protein.
SQ SEQUENCE 517 AA; 58502 MW; 4FCF5CD5C2B5AE0A CRC64;

Query Match 4.7%; Score 99; DB 1: Length 517;
Best Local Similarity 18.0%; Pred. No. 1.3;
Matches 44; Conservative 35; Mismatches 65; Indels 100; Gaps 8;

QY 35 KNGKGGVLP-----KEKFNKPPSTPRAYNR-----EOKELNR 68
Db 263 RGGKTYGDDIPACQIDEGSKMKAFIIPPNSSQAQYNLGNMYPFTVSGSLVSSDAQLFNR 322
QY 69 -----WYNPIILNRVANOTGELATSPNTSHLSYCEPDSVTMTAVTDFNNLP 113
Db 323 PFWLQAOQHNGHNGICWFNLQFVTVVDNT-----RNTNFSISINSDGTDYSKITDYS-- 374
```

```
QY 114 DRFKDFLLYLRCRNYSLLIDQPKKCAKPFELLALIAKSLIPFARROAIRSNGRETNVGN 173
Db 375 QKFTYLRHVVEYELSLLQLCKRVPKAEIL-----AQINAMN 412
QY 174 QTVVYVFLGKTPPEONHHPDLSMDLKFESDKHODILMMNYRDTFFNLSLKEVFLRWYST 233
Db 413 SNILEEWQLGFGVPAPDN-----SIQDTY-----RYIDSLAT 443
QY 234 SCPD 237
Db 444 RCPD 447
```

Search completed: October 20, 2003, 23:02:30
Job time : 13.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:56:05 ; Search time 18.5 Seconds
(without alignments)
2063.728 Million cell updates/sec

Title: US-09-804-357b-2

Perfect score: 2129
Sequence: 1 MSVGRRRVKKLGLIMANVF.....RKPOEMIDWSQLSPNLKC 397

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	11.2	739	Tl0648	hypothetical prote
2	218	10.2	507	Tl8639	hypothetical prote
3	193	9.1	657	Tl8639	protein Tl8639.13 [i
4	192.5	9.0	642	D96777	hypothetical prote
5	186.5	8.8	384	T20446	hypothetical prote
6	185.5	8.7	262	T20203	hypothetical prote
7	183	8.6	345	T20031	hypothetical prote
8	181	8.5	357	T20029	hypothetical prote
9	181	8.5	376	T20028	hypothetical prote
10	164.5	7.7	332	T20028	hypothetical prote
11	164.5	7.7	684	T86394	protein T24p13.20
12	157	7.4	325	T24762	hypothetical prote
13	155	7.3	376	T24925	hypothetical prote
14	153	7.2	338	T24743	hypothetical prote
15	151	7.1	395	T86458	probable elicitor
16	150.5	7.1	399	T86453	avrs homolog F9L11
17	140	6.6	398	T86353	protein F2E2.6 [im
18	135.5	6.4	368	T15096	hypothetical prote
19	130.5	6.1	348	T31918	hypothetical prote
20	130.5	6.1	359	T87937	protein F14B6.6 [i
21	130.5	6.1	385	T20879	hypothetical prote
22	128	6.0	401	T86251	hypothetical prote
23	126.5	5.9	406	T04817	hypothetical prote
24	123	5.8	455	T29555	hypothetical prote
25	122.5	5.8	279	T20443	hypothetical prote
26	122.5	5.8	548	T86456	unknown protein [i
27	120.5	5.7	334	T32256	hypothetical prote
28	118	5.5	349	T24744	hypothetical prote
29	116.5	5.5	404	T86186	hypothetical prote

30	113.5	5.3	409	2	A84733	hypothetical prote
31	113.5	5.3	479	2	D86187	hypothetical prote
32	112.5	5.3	414	2	B96808	protein F28K19.2 [i
33	112.5	5.3	560	2	A86214	hypothetical prote
34	112	5.3	353	2	C96573	protein F12M16.19
35	112	5.3	753	2	T24745	hypothetical prote
36	109.5	5.1	253	2	T22387	hypothetical prote
37	104.5	4.9	1026	2	A49750	beta-galactosidase
38	104	4.9	885	2	B70393	hypothetical prote
39	100	4.7	432	2	D72255	beta-fructofuranos
40	99	4.7	374	2	G88799	protein T04B2.3 [i
41	99	4.7	412	2	T24441	hypothetical prote
42	99	4.7	440	1	S33479	ferredoxin-NADP re
43	99	4.7	517	2	S36496	LI protein - human
44	98	4.6	1018	2	E64451	type I restriction
45	97.5	4.6	610	2	S59394	protein kinase RCK

ALIGNMENTS

RESULT 1

Tl0648
hypothetical protein Tl3K14.220 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C:Accession: Tl0648
R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Meves, H.W.; Mayer, K.F.X.; Le
submitted to the Protein Sequence Database, June 1999
A:Reference number: 216991
A:Accession: Tl0648
A:Molecule type: DNA
A:Residues: 1-739 <BEV>
A:Cross-references: EMBL:AL080282; GSPDB:GN00062; ATSP:Tl3K14.220
A:Experimental source: cultivar Columbia; BAC clone Tl3K14
C:Genetics:
A:Gene: ATSP:Tl3K14.220
A:Map position: 4
A:Introns: 59/3; 96/1; 358/1; 438/3; 535/3; 578/3; 634/3; 670/3

Query Match 11.2% Score 238; DB 2; Length 739;
Best Local Similarity 22.8%; Pred. No. 1.5e-11;
Matches 75; Conservative 65; Mismatches 133; Indels 56; Gaps 10;

Qy	81	TGEL-----ATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFXDFLLYLRCRNYSLIDQ	134
Db	451	TGVDIHSIHATSUSTSHPSFSPKATIEFSEWAPPLPG-----	490
Qy	135	PKCAKKPF-LLLAIKSLIPFAFRRQAIRESWGRETWGNQTVVRFVFLGKTPPEONHPD	193
Db	491	-----TFPRLFMGVLSATNHFSEMAVRKTMQHPHSIKSSDVVARFFVALNP-----RKE	540
Qy	194	LSMDLKFESDKHQDILMNNYRDTFFNLKLEFLRWVSTSCDPAEFVFGDDDFVFNTH	253
Db	541	VNMLKKEAEYFGDIVLPEDRYELVLTATICEFGNVTAP---YIMKCDDDTIFRVE	597
Qy	254	HLNVLNSLSKAKDLFGIDVHNAGPHRDKLKYIPEVETGVVYPPYAGGGGFLYSG	313
Db	598	SILKQIDGVSPSEKS--LYFGNGLNLRHRLPTGKWTVTWEE-WPEAVYPPYANGFGYIIS	654
Qy	314	PALLRLYSATSR--VHLYPIDDVTYTGMLQKLGIVPEKHKGFTFDIEENKKNICSY--	369
Db	655	NTAKYIVSQNSRHKLRLFKMEDVSGMLVEQFNASMQ-----PVEYSHSNKFCQYGC	706
Qy	370	-IDLMLVHSRRKPOEMIDWSQLSPNLKC	397
Db	707	TILNYTAHYQSPSQMCLMDNLKGRPOC	735

RESULT 2

Tl8639
hypothetical protein B0024.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans

```
C>Date: 15-Oct-1999 *sequence_revision 15-Oct-1999 *text_change 15-Oct-1999
C/Accession: T18639
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: 219001
A:Accession: T18639
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-507 <MIL>
A:Cross-references: EMBL:271178; PIDN:CAA94876.1; GSPDB:GN00023; CESP:B0024.3
A:Experimental source: clone B0024
C:Genetics:
A:Gene: CESP:B0024.3
A:Map position: 5
A:Introns: 69/2; 118/2; 143/2; 181/1; 246/2; 281/3; 393/2; 437/3

Query Match      10.2%; Score 218; DB 2; Length 507;
Best Local Similarity 22.8%; Pred. No. 4e-10;
Matches 92; Conservative 70; Mismatches 156; Indels 86; Gaps 15;

QY 58 YWNR--OEKLNRYNPILNR-----VANQTGELAT 86
DB 125 YWDSRYAQEKISPTFFVMSHLPMDLPDLYFLDLDQGLIIVECTVVIILENTYSNVT 184
QY 87 SPNTSHLSYCEPDSTVMTATDFNNLPDRKD-----FLLYLRGNYSYL-LIDOPKK-C 138
DB 185 FLNLDRIYSYDAPFEVNTSFSSKSHLSKSGSTNOFYHAFKDNHNYQIITYPKKQC 244
QY 139 AKKPELLLAIKSLIPFARQAIRESWGRETN---VGNQTVVRVFLGKTTPPDNHPDLS 195
DB 245 SNNTKQITLSTAGNFDIRQAIRETWAMPNNSHIVANNDRVIFSIKTSNE---FLN 300
QY 196 DMLKFESDKHODILMNYRDIIFENLSKEVLFLRWVSTSCPDPAEFVKGDODVFNTHI 255
DB 301 FALQKEIFEDDDIMVDLYSEELLKVAHLSKYQSHCOLADQLKIDDDMAYDMGGS 360
QY 256 LNVLSLSKSKAKDLFGDVIHAGPHRDKKLKYYIPEVFT-GVYPPVAGGGGFLYSGP 314
DB 361 YRSLEDKKQASINGI-SGIWNKSPVPRKKHRYVPKLYSEKPEPPYIDGPVILGKN 419
QY 315 ALLRLYSATSRYHLYPIDDYV-TGMCLOKL-----GLVPEKHKGFRFTD 357
DB 420 AVPRMLEAKNYNOMWIEDVFTGVTGKALKIKQINWANHLLRYVELIPSLK----- 473
QY 358 IEKKNKNCYSIDLML--VHSRK-PQEMIDIWSLOSPNLKC 397
DB 474 -----CSKGGVPLIYAVHNKGFQMIHGYQKLR--GVRK 506

RESULT 3
G86397
protein T7N9.18 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 *sequence_revision 02-Mar-2001 *text_change 31-Mar-2001
C/Accession: G86397
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G86397
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-657 <STO>
A:Cross-references: GB:AE005172; NID:g8778958; PIDN:AAF79857.1; GSPDB:GN00141
C:Genetics:
A:Gene: T7N9.18
```

A:Map position: 1

```
Query Match      9.1%; Score 193; DB 2; Length 657;
Best Local Similarity 26.7%; Pred. No. 6.7e-08;
Matches 67; Conservative 43; Mismatches 95; Indels 46; Gaps 11;

QY 158 QAIRESWGRETNVGNQTVVRVFLGKTTPPDNHPDLSDKLPESDKHODILMNYRDTF 217
DB 425 RMAYRKSMMQOKLVRSSKVVAREFFVALHARKEVND---LKKEAEYFGDIVIPYMDHY 480
QY 218 FNLSLKEVLFLRW-VSTSCPDPAEFVKGDODVFNTHHLLNLYLSLSKSKAND-LFIDGY 275
DB 481 DLVVLKTVAICEYGVNTVA--AKYVMKCDODTTFVRDVA---IQEAEKVGKRESLYIGNI 535
QY 276 IHNAGPHRDKKLKYYI---PEVFTGYVPPYAGGGGFLYSGPALLRLYSAT----- 323
DB 536 MFNHHPLRTCKWAVTPEWPEEY---YPPYANGPGVILS-----YDVKAFIVDDFEQ 584
QY 324 SRVULYPIDOVYTGMCLOKGLVPEKHKGFRFTDIEKKNKNCISY---IDLMLVHSRXP 380
DB 585 KRLSLFKHEDVSGMMWVEKEN-----ETRPVAVVHSLAKFCQFCIEDYFTAHYQSP 635
QY 381 QEMIDIWSLOQ 391
DB 536 ROMICMWDKLO 646

RESULT 4
D96777
hypothetical protein F25A4.23 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 *sequence_revision 02-Mar-2001 *text_change 31-Mar-2001
C/Accession: D96777
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alo,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96777
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-642 <STO>
A:Cross-references: GB:AE005173; NID:g5882743; PIDN:AAD55296.1; GSPDB:GN00141
C:Genetics:
A:Gene: F25A4.23
A:Map position: 1

Query Match      9.0%; Score 192.5; DB 2; Length 642;
Best Local Similarity 24.0%; Pred. No. 7.1e-08;
Matches 62; Conservative 57; Mismatches 114; Indels 25; Gaps 8;

QY 144 LLLAKSLIPHFARQAIRESWGRETNVGNQTVVRVFLGKTTPPDNHPDLSDKLPESD 203
DB 395 IFICILSAGNHFSERMAVRKSNQHVLIITSAKVAREFFVAL-----HGRKEVNVELKKEAE 450
QY 204 KHODILMNYRDTFFNLNLSKEVLFLRWVSTSCPDPAEFVKGDODVFNTHHLLNLYLSLS 263
DB 451 YFGDIVLVPYMDSYDLVWLKTVAICEHGALEF-SAKYIMKCDODTTFVLGAVINEVKVP 509
QY 264 KSAKOLFIQDGVTHNAGPHRDKK--LKYYIPEVFTGYTPPYAGGGGFLYSGPALLRLYS 321
DB 510 EGRS--LYIGNMNYHKKPLRGKWAFTY---EWPEDYPPPYANGPGVILSSDIARFIYD 564
QY 322 ATSR--VHLVPIDDVYTGMCLOKGLVPEKHKGFRFTDIEKKNKNCISY---IDLMLVH 376
DB 565 KFERHKLRLFKMEDVSVGMWVE-----HFKNTNPVDYRHSRFRFCQCGCVENYTTAH 616
```

```
QY 377 SRKQEMIDIMSOLOSPN 394
Db 617 YQSPRQMICLMDKLRGN 634

RESULT 5
T20446
hypothetical protein E03H4.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: T20446
R:Dobson, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19276
A:Accession: T20446
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-384 <WIL>
A:Cross-references: EMBL:Z81492; PIDN:CAB04032.1; GSPDB:GN00019; CESP:E03H4.11
A:Experimental source: clone E03H4
C:Genetics:
A:Gene: CESP:E03H4.11
A:Map position: 1
A:Introns: 46/1; 76/1; 196/3; 263/2; 305/3; 335/3
C:Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

Query Match 8.8%; Score 186.5; DB 2; Length 384;
Best Local Similarity 20.2%; Pred. No. 1.1e-07;
Matches 79; Conservative 82; Mismatches 133; Indels 97; Gaps 15;

QY 1 MSVGRRRVKLLGILMANVFIYLIVEYKNSQDKNGKGGVLPKPKFWPPPTPRAYWN 60
Db 32 LDLGRSPKLPKIVRVASL-----RRNRSSFHSE-----IPKE-----NCTRSGWL 74

QY 61 REQELKRWNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRKDFL 120
Db 75 KD-----STAPETQDFG-----SEFVISFADIQ----- 97

QY 121 LYLRCRNYSLIDQPKKCAKPPLLAIAKSLIPHFARRQAIRRESWGREN---VGNQTV 176
Db 98 -----KNYTLWLPYEIEASOEKDIIMVASRTDSYARRNIMROTWNKSDSEIVANGRM 152

QY 177 VRFLKCTPPEDNHPDLSMDLKPFESDKHODILMWNRYDTFFNLKKEVLFLRWVYSTSCP 236
Db 153 KPLFLVGLTGDYK---MKKVMQEAKEYGLDIIVVDMNDYEELTYKSLAILLYGVSKAP 209

QY 237 DAEVFKGDDVDVFNTHHILNLSLSKSAKADLFIGDVIRHAGP-----H 282
Db 210 RYQMGIKIDEDVI-----FFPQKLTALYEQGII-DATPLCAYGVKIQAGARIF 256

QY 283 RDKKLKYYIPEVFT-GVYPPYAGGGFLYSGPALLRLYSATSRVHLYPIDDY-TGMCL 340
Db 257 RDKNDRWIVPESSYSCSKFPEYVSGMLVYTWEEAAQOIISTKYRDFIQVEDVFLTGILA 316

QY 341 QKGLGVLPKPKHGRFTF--DIEKKNKNIYCSY 369
Db 317 EDLGISVRLNPKFYKPNIDESKSVDIIAW 347

RESULT 6
T20203
hypothetical protein C54C8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20203
R:Dobson, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19237
A:Accession: T20203
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-262 <WIL>
A:Cross-references: EMBL:Z83102; PIDN:CAB05465.1; GSPDB:GN00019; CESP:C54C8.3

Query Match 8.6%; Score 183; DB 2; Length 345;
Best Local Similarity 22.5%; Pred. No. 1.9e-07;
Matches 83; Conservative 71; Mismatches 137; Indels 78; Gaps 16;

QY 48 FKKPPSTPRAYWNREQELNRWNPILNRVANOTIGELATSPNTSHLSYCEPDSTVMTAVT 107
Db 24 FKKLPQNPKSSSVLEERHCYQSGWN--ISTLSRPNIDFGSS-----FIV 64

QY 108 DFNNLPDRKDFLILYLCRNYSLLIDQPKKCAKPPLLAIAKSLIPHFARRQAIRRESWG 167
Db 65 SFANIHSQK--WYIL-----PKMKNVTKHDIILMLVVSCKTKNFARRNVLRSWTWN 113

A:Experimental source: clone C54C8
C:Genetics:
A:Gene: CESP:C54C8.3
A:Map position: 1
A:Introns: 19/2; 117/3; 184/2; 226/3

Query Match 8.7%; Score 185.5; DB 2; Length 262;
Best Local Similarity 23.6%; Pred. No. 8.1e-08;
Matches 61; Conservative 58; Mismatches 95; Indels 45; Gaps 10;

QY 127 NYS-LLIDQPKKCAKPPLLAIAKSLIPHFARRQAIRRESWGREN---VGNQTVRVVELL 182
Db 20 NYSWLYLPEIETSQEKOILIVASRTDSYARRNIIROTWMKSKANSEIVANGMKPFLFV 79

QY 183 GKTTPEDNHPDLSMDLKPFESDKHODILMWNRYDTFFNLKKEVLFLRWVYSTSCPDAEFV 242
Db 80 GLTPGEYK---MKKVMQEAKEYGLDIIVVDMNDYEELTYKSLAILLYGVSKAPRYQMIG 136

QY 243 KGDDOV-----FVNTHHILNLSLSKSAKADLFIGDVIRHAGPDRKDKLK 288
Db 137 KIDEDVMFFPKLTLYDQGFIDATPLRIY--GLKMQSGANIF-----RDKTHR 183

QY 289 YVYIPEVFT-GVYPPYAGGGFLYSGPALLRLYSATSRVHLYPIDDY-TGMCLQKGL- 345
Db 184 WYVPESSYSCSKFPEYVSGMLVYTWEEAAQOIISTKYRDFIQVEDVFLTGILAEOLGIS 243

QY 346 ---VPEKHK-----GFRTF 356
Db 244 VKNLPEFYKYPDVGFFSF 262

RESULT 7
T20031
hypothetical protein C47F8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T20031; T22806
R:McLay, K.
submitted to the EMBL Data Library, December 1997
A:Reference number: Z19212
A:Accession: T20031
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-345 <WIL>
A:Cross-references: EMBL:AL009246; PIDN:CAAL15841.1; GSPDB:GN00019; CESP:C47F8.3
A:Experimental source: clone C47F8
R:Kershaw, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19617
A:Accession: T22806
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-345 <WIL>
A:Cross-references: EMBL:Z81553; PIDN:CAB04502.1; GSPDB:GN00019; CESP:C47F8.3
A:Experimental source: clone F56H6
C:Genetics:
A:Gene: CESP:C47F8.3
A:Map position: 1
A:Introns: 23/1; 49/3; 169/3; 236/2; 278/3; 308/3
C:Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

Query Match 8.6%; Score 183; DB 2; Length 345;
Best Local Similarity 22.5%; Pred. No. 1.9e-07;
Matches 83; Conservative 71; Mismatches 137; Indels 78; Gaps 16;

QY 48 FKKPPSTPRAYWNREQELNRWNPILNRVANOTIGELATSPNTSHLSYCEPDSTVMTAVT 107
Db 24 FKKLPQNPKSSSVLEERHCYQSGWN--ISTLSRPNIDFGSS-----FIV 64

QY 108 DFNNLPDRKDFLILYLCRNYSLLIDQPKKCAKPPLLAIAKSLIPHFARRQAIRRESWG 167
Db 65 SFANIHSQK--WYIL-----PKMKNVTKHDIILMLVVSCKTKNFARRNVLRSWTWN 113
```


QY 254 HILYNLSKSKARDLFGDVIHNAHPDRKK-LKYYIPEVY--IGVYPPYAGGGGL 310
 Db 177 KLIPULDDCKVIDPDAARFYGOLLKEGEVIRKKDAHWYVPDYAYNCTG-YPAYVAGPFYL 235
 QY 311 YSGPALLRLYSATSRVHLYPIDD-VYTGMCLOKLGVLPEK---HKGFRFTDIEEKNKNI 366
 Db 236 ATRKAAKLVLFKTFKQFNPMTVEDSLITGILANDLG-IPRKNLEHVRHYRDIQDNESKEI 294
 QY 367 CSYIDLMLVHSRK 379
 Db 295 LAW-----HSFK 301

RESULT 11
 F86394
 protein T24P13.20 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: F86394
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, I.H.; Dewar, K.;
 Ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: F86394
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-684 <STO>
 A:CROSS-references: GB:A005172; NID:g9295733; PIDN:AAF87039.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T24P13.20
 A:Map position: 1

Query Match 7.7%; Score 164.5; DB 2; Length 684;
 Best Local Similarity 22.1%; Pred. No. 1.6e-05;
 Matches 64; Conservative 53; Mismatches 102; Indels 71; Gaps 13;

QY 144 LLLATKSLIPHFAARQAIRESGRETNV-GNOTVVRVFL-GKTPPEDNHFDLSMLKFE 201
 Db 413 LVIGVFSTANFKRMARVETWQTDVRSRVARVFFVGLHKSP-----LVNLE 462
 QY 202 SDKHODILMNVRDTFFNLSLKEVL-----FLRWVS-----TS 234
 Db 463 -----LWNEARTYGDVQLMPFVDYVSLISWKTALICIFGLSLYCFASITSYFKQTE 514
 QY 235 CPDAEFVFGDDVDVFNTHILNLYNLSKSKAKDLFGDVIHNAHPDRKKLYIPEV 294
 Db 515 VDSAKFIMKTDDAFVDEVLSSMTNNTRG---LIYGLINSQSOPIRNPDSDKWIYSY 572
 QY 295 FY-----TGVPYPPYAGGGFLYS---GPALLRLYSATSRVHLYPIDDVTGMCLOKL- 343
 Db 573 FLLKKQEMPEEKYPWAGPGVINSRDIAESVGKLFK-EGNLKNEKLEDVAMGWIABELT 631
 QY 344 --GLVPEKHGKGRFTDIEEKNKNCISYIDLMLVHSRKPQEMIDVWSOLO 391
 Db 632 KHGLEPHVENDGRIISDGCKD-----GVV---VAHYQSPAEMTCLWRKYQ 673

RESULT 12
 T24762
 hypothetical protein T09F5.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T24762
 R:Morimore, B.
 submitted to the EMBL Data Library, December 1996
 A:Reference number: Z19933

A:Accession: T24762
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-325 <WIL>
 A:CROSS-references: EMBL:Z83239; PIDN:CAB05807.1; GSPDB:GN00023; CESP:T09F5.1
 A:Experimental source: clone T09F5
 C:Genetics:
 A:Gene: CESP:T09F5.1
 A:Map position: 5
 A:Introns: 19/1; 42/1; 155/3; 223/2; 265/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

Query Match 7.4%; Score 157; DB 2; Length 325;
 Best Local Similarity 23.4%; Pred. No. 2.5e-05;
 Matches 64; Conservative 47; Mismatches 102; Indels 60; Gaps 11;

QY 100 STYMTAVTDF-NLIPDRFKDF-----LLYLRCNRYSLIDOPKKCAKKPPELLLAIKSLIP 153
 Db 42 STLNSRDRGSHYEISFADIQGSFEWLYL-----PKFELNNPELLIATSRPD 90
 QY 154 HFARRQAIRESW-GRETNVGNQTVVRVFLGKTPPEDNHFDLSMLKFESDKHQDILMN 212
 Db 91 DESRRNAIRKTMNOKTN-----QITSEFMYGLSSKTD--KVRDIVMREAEIYRDIWVTS 144
 QY 213 YRDTFFNLSLKEVFLRWVSTSCPDAAEFVEKGGDDV-----FVNTHIL 256
 Db 145 LEDSYTKLAPKTLISILLYAVSKVPSAQLIGRVDGDLVFFPNLFQSLDKDNYFINTNNS 204
 QY 257 NYLNSLSKSKAKDLFGDVIHNAHPDRKKLK---YYIPEVFTGVVYPPYAGGGFLYS 313
 Db 205 IY-----GYIAEKGKPTTSCKCKSRNLF--FKCSNYLSLSGPFELLTR 248
 QY 314 PALLRLYSATSRVHLYPIDD-VYTGMCLOKLG 345
 Db 249 PAAEKLNLNASKHRDFHOIDDQLITGOMADDAGV 281

RESULT 13
 T24925
 hypothetical protein T15D6.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T24925
 R:Dobson, R.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19956
 A:Accession: T24925
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-376 <WIL>
 A:CROSS-references: EMBL:Z83125; PIDN:CAB05616.1; GSPDB:GN00019; CESP:T15D6.5
 A:Experimental source: clone T15D6
 C:Genetics:
 A:Gene: CESP:T15D6.5
 A:Map position: 1
 A:Introns: 59/1; 87/1; 207/3; 274/2; 316/3; 339/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

Query Match 7.3%; Score 155; DB 2; Length 376;
 Best Local Similarity 21.5%; Pred. No. 4.4e-05;
 Matches 58; Conservative 56; Mismatches 100; Indels 56; Gaps 11;

QY 144 LLLAIAKSLIPHFAARQAIRESGRETNVGNQTVVR-----VFLGKTPPEDNHFDLSM 197
 Db 128 ILMIVASRPGSVSRKKVLRKTW---MNKANSKIIRNGRMQVLELVGMVAGD---RDLMA 181
 QY 198 LKTESDKHODILMNVRDTFFNLSLKEVFLRWVSTSCPDAAEFVEKGGDDVVF----- 250
 Db 182 VKAEASFGDIIVNNLEDYDNLNPFKVLISLLLVGTNKAADFKIIGKIDDDVFPPDRLRTP 241
 QY 251 -----NTHILNLYNLSKSKAKDLFGDVIHNAHPDRKKLYIPEVFTGV-Y 300
 Db 242 LLDENVIDSSSYSIYGL-----SQDELVY-----RNETKPMVYVPEYATYNTKY 286

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 23:00:41 : Search time 12.5 Seconds
(without alignments)
1343.793 Million cell updates/sec

Title: US-09-804-357B-2

Perfect score: 2129

Sequence: 1 MSVGRRRVKLLGLIMANVF.....RKPEMIDWSQLQSPNLKC 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	635.5	29.8	378	4	US-09-482-180A-2
2	578	27.1	397	4	US-09-459-133-2
3	565	26.5	389	4	US-09-459-133-13
4	480.5	22.6	326	2	US-09-055-097-6
5	347.5	16.3	325	2	US-09-055-097-5
6	308.5	14.5	331	4	US-09-996-243-209
7	287	13.5	378	2	US-09-055-097-1
8	142	6.7	372	1	US-08-207-904-10
9	103	4.8	363	4	US-09-464-035A-1
10	103	4.8	363	4	US-09-849-562A-1
11	103	4.8	363	4	US-09-849-031A-1
12	101	4.7	350	4	US-09-464-035A-11
13	99.5	4.7	363	4	US-09-464-035A-3
14	97	4.6	363	4	US-09-464-035A-5
15	88	4.1	624	2	US-08-756-317-9
16	87.5	4.1	1024	3	US-09-091-117-5
17	84	3.9	379	4	US-09-328-352-5405
18	83	3.9	476	4	US-09-134-001C-4117
19	82	3.9	409	3	US-09-023-809B-3
20	81.5	3.8	912	2	US-08-951-871-2
21	81.5	3.8	2165	1	US-08-514-975B-2
22	81.5	3.8	2165	5	PCT-US95-12507-2
23	80.5	3.8	990	1	US-08-232-540-2
24	80.5	3.8	990	1	US-08-428-949A-2
25	80.5	3.8	990	1	US-08-428-948A-2
26	80.5	3.8	990	2	US-08-428-946-2
27	80.5	3.8	990	5	PCT-US95-04656-2

28	80.5	3.8	1013	1	US-08-233-008A-8	Sequence 8, Appli
29	80	3.8	784	4	US-09-982-008B-23	Sequence 23, Appli
30	80	3.8	1138	1	US-07-973-320-2	Sequence 2, Appli
31	80	3.8	1138	1	US-07-973-320-4	Sequence 4, Appli
32	79.5	3.7	320	4	US-09-134-001C-4096	Sequence 4096, Ap
33	79.5	3.7	389	4	US-09-464-035A-7	Sequence 7, Appli
34	79.5	3.7	410	1	US-08-792-283A-8	Sequence 8, Appli
35	79.5	3.7	410	2	US-09-105-908-8	Sequence 8, Appli
36	79.5	3.7	410	3	US-09-271-713-8	Sequence 8, Appli
37	79.5	3.7	610	3	US-08-484-661A-33	Sequence 33, Appli
38	79.5	3.7	610	3	US-08-484-661A-35	Sequence 35, Appli
39	79.5	3.7	610	3	US-08-656-664-33	Sequence 33, Appli
40	79.5	3.7	610	3	US-08-656-664-35	Sequence 35, Appli
41	79.5	3.7	610	3	US-08-656-664-54	Sequence 54, Appli
42	79.5	3.7	610	5	PCT-US96-09641-33	Sequence 33, Appli
43	79.5	3.7	610	5	PCT-US96-09641-35	Sequence 35, Appli
44	79.5	3.7	610	5	PCT-US96-09641-54	Sequence 54, Appli
45	79.5	3.7	906	1	US-08-220-151-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1

US-09-482-180A-2
; Sequence 2, Application US/09482180A
; Patent No. 6361985
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Yamamoto, Gayle
; APPLICANT: Gao, Zeren
; APPLICANT: Whitmore, Theodore E.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING
; FILE REFERENCE: 98-80
; CURRENT APPLICATION NUMBER: US/09/482.180A
; CURRENT FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/115,721
; PRIOR FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-482-180A-2

Query Match	29.8%	Score 635.5;	DB 4;	Length 378;
Best Local Similarity	39.5%;	Pred. No. 2.1e-60;		
Matches	146;	Conservative 52;	Mismatches 117;	Indels 55; Gaps 11;
Qy	40	GVIIPEKEFWKPPSTPRA---YWNREQEKLNRWNPILNRVANOTGELATSPNTSHLSYC	96	
Db	44	GCLEFLRAAKPAGDPTAHQPFW-----APPIPRHSC	76	
Qy	97	EPDSVTMTAVTDENLPDRKDFLLYLCRNYSLLIDOPKKCAKKPFLLLAIAKSLIPHFA	156	
Db	77	PNHVTSSASL---SLPSRHLFLTYRHCNRFSLIL-EPSCSKDTELLLAIAKSPGHVE	132	
Qy	157	RRQAIRESWGRETNRVG----NOTVRVFLL---GKTPPEDNHPDLSMDLKPFESDKHDI	209	
Db	133	RAAAIRSTWGR---VGGWARGRLKLVLLGVAGSAPP-----AQLLAYESREFDDI	182	
Qy	210	MNMYRDTFNLKLEVLFLRWVTSCTPDAEFVFGDDDFVNTHTHILNLSLSKSKAKD	269	
Db	183	QWDFTEDEFNLTKELHLQRWVAACPAQHFMKLGDDDFVHVNPVLEFLDGM--DPAQD	240	
Qy	270	LFIGDVHINAGPHRDKLKYIPEVFTGV-YPPVAGGGGFLYSGPALLRLYSATSRVHL	328	
Db	241	LVUGDVIOALPNRTKVKYFIPPSMYRATHYPPYAGGGGVMSRATVRRLQALMEDAEL	300	
Qy	329	YPIDDDVTGMCLOKLGVLVPEKHGKGRFTDIEEK-NKKNICSYIDLMLVHSRKRPOEMIDW	387	


```

; FILING DATE: Filed Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Certone, Michael C.
; REGISTRATION NUMBER: 39,112
; REFERENCE/DOCKET NUMBER: PF-0490 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 326 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 2745735
; US-09-055-097-6

Query Match 22.6%; Score 480.5; DB 2; Length 326;
Best Local Similarity 33.7%; Pred. No. 1.1e-43;
Matches 112; Conservative 64; Mismatches 127; Indels 29; Gaps 9;

QY 69 WYNFILNRVAMOTGELATSPNTSHLSYCEPDSVTMTAVTDNPNLPORFKDFLLYLRCRNY 128
DB 21 WYLSITRPTSYTG----SKPFSLTVARKNFT-----FGNIRTR-----PINPHSF 63

QY 129 SLLIDOPKKCAK--PFLLLAIKSLIPHFAAROAIRESWGRETNGVGTVVYRVFLGKTPP 187
DB 64 EFLINEPKCEKNPFLVLIISTHKEFDARQALRETWGDENPKGIKATLELLGK--- 120

QY 188 EDNHPDLSMLKFESDKHODILMNYRDTFFNLSLKEVLFRLWVYSTCSDAEVFVKGDOD 247
DB 121 -NADPVLNQMYEQESQIFHDIIEDFDSYHNLTLTLMCMRWVATFCSKAKYVMKTDSD 179

QY 248 VFVNTHLNLYNLSLSKAKDLFIGDVIIHNAHPHROKKLKYIPEVY--TCGYPPYAGG 306
DB 180 IFVNMDLNLYKLPSTKPRRYFTGYVI--NGGPIRDVRSKWMYPRDLYPDSNYPFCSG 238

QY 307 GGFLYSGPALLRLYSATSRVHLYPIDOVYTGMCLOKGLYPEKHKGFRTDIEKNKNI 366
DB 239 TGYIFSADVRLIYKTSHTFLLHEDVYVGLCLRLKLIHPFQNSGFNHW-----KMAYSL 294

QY 367 CSYIDLMLVHSRKPEMIDIMSOLOS--PNLKC 397
DB 295 CRYRRVITVHQISPEMHRWINDMSSKKHLRC 326

RESULT 5
US-09-055-097-5
; Sequence 5, Application US/09055097
; Patent No. 595282
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Kari J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; FILING DATE: Filed Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Certone, Michael C.
; REGISTRATION NUMBER: 39,112
; REFERENCE/DOCKET NUMBER: PF-0490 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 326 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 2745735
; US-09-055-097-6

Query Match 22.6%; Score 480.5; DB 2; Length 326;
Best Local Similarity 33.7%; Pred. No. 1.1e-43;
Matches 112; Conservative 64; Mismatches 127; Indels 29; Gaps 9;

QY 69 WYNFILNRVAMOTGELATSPNTSHLSYCEPDSVTMTAVTDNPNLPORFKDFLLYLRCRNY 128
DB 21 WYLSITRPTSYTG----SKPFSLTVARKNFT-----FGNIRTR-----PINPHSF 63

QY 129 SLLIDOPKKCAK--PFLLLAIKSLIPHFAAROAIRESWGRETNGVGTVVYRVFLGKTPP 187
DB 64 EFLINEPKCEKNPFLVLIISTHKEFDARQALRETWGDENPKGIKATLELLGK--- 120

QY 188 EDNHPDLSMLKFESDKHODILMNYRDTFFNLSLKEVLFRLWVYSTCSDAEVFVKGDOD 247
DB 121 -NADPVLNQMYEQESQIFHDIIEDFDSYHNLTLTLMCMRWVATFCSKAKYVMKTDSD 179

QY 248 VFVNTHLNLYNLSLSKAKDLFIGDVIIHNAHPHROKKLKYIPEVY--TCGYPPYAGG 306
DB 180 IFVNMDLNLYKLPSTKPRRYFTGYVI--NGGPIRDVRSKWMYPRDLYPDSNYPFCSG 238

QY 307 GGFLYSGPALLRLYSATSRVHLYPIDOVYTGMCLOKGLYPEKHKGFRTDIEKNKNI 366
DB 239 TGYIFSADVRLIYKTSHTFLLHEDVYVGLCLRLKLIHPFQNSGFNHW-----KMAYSL 294

QY 367 CSYIDLMLVHSRKPEMIDIMSOLOS--PNLKC 397
DB 295 CRYRRVITVHQISPEMHRWINDMSSKKHLRC 326

RESULT 5
US-09-055-097-5
; Sequence 5, Application US/09055097
; Patent No. 595282
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Kari J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; FILING DATE: Filed Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Certone, Michael C.
; REGISTRATION NUMBER: 39,112
; REFERENCE/DOCKET NUMBER: PF-0490 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1150971
; US-09-055-097-5

Query Match 16.3%; Score 347.5; DB 2; Length 325;
Best Local Similarity 32.0%; Pred. No. 3.2e-29;
Matches 90; Conservative 53; Mismatches 119; Indels 19; Gaps 7;

QY 118 DELLYLRCRNY--LLIDOPKKCAKPPFLLLAIKSLIPHFAAROAIRESWGRETNGVGT 175
DB 58 DKFAYLRVPTAEAYPVDPQPAR-----LTMLIKSAGVNSRRREAIRRTWGYEGRFSDVH 111

QY 176 VYRVFLGKTPEDNHPDLSMLKFESDKHODILMNYRDTFFNLSLKEVLFRLWVYSTC 235
DB 112 LRRVFLGTA--EDSEKDA----WESREHGOIQADTFDAYFNILATMLGNRWASEQF 165

QY 236 PDAAEFVFKGDDOVFYNTTHILNLYNLSLSKAKDLFIGDVIIHNAHPHROKKLKYIIP-EV 294
DB 166 NRSEYFLVDDDDYYSAKNVLKFLGRGQSHOPELLFAGHVQIISPLRHKFSKWVSL 225

QY 295 FYTGVPYVYAGGGFLYSGPALLRLYSATSRVHLYPIDOVYTGMCLOKGLVPEKHKGFR 354
DB 226 YPFDRMPYVTAGAFILSQALRQLYAAVHPLPFRFDDVYLGIVALKAGISLQRCDDFR 285

QY 355 TFDIEKNKNKICSYIDLMLVHS--RKPOEMIDIMSOLOSPN 394
DB 286 ---FHRPAYKGPDSYSSVSIASHEFGDPEENTRVWNECRSAN 323

RESULT 6
US-09-996-243-209
; Sequence 209, Application US/09996243
; Patent No. 6478825
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavir, Ivar J.
; APPLICANT: Napier, Mary A.
```

APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: ROY, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zhenin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE INVENTION: Acids Encoding the Same
FILE REFERENCE: P2730PLC13
CURRENT APPLICATION NUMBER: US/09/996,243
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/0889105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24

```

; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match      14.5%; Score 308.5; DB 4; Length 331;
Best Local Similarity 30.2%; Pred. No. 5.7e-25;
Matches 84; Conservative 56; Mismatches 123; Indels 15; Gaps 8;

QY 126 RNYSLIDQPKKCA--KKPFLLLAIKSLIPHFARQAIRESMGRETNGVTVVVFLLG- 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 QDFHTEHNSCHQNFELVILVTSHPDSVKARQAIRVWGEKSKWGVEVLTFFLLGQ 120

QY 184 KIPPDNHPDLSMLKFESDKHQDILMMNYRDTFENLSLKEVLFLRWVYSTCPDAEFVK 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 EAEKEDKMLALS--LEDEHLLYGDIIRODFLDYNNLTKTINAFRWVTEFCPNKYVMK 178

QY 244 GDDDFVFNTHHLYNLSLSKSKAKDLFGDVIHAGPHRDKLKYIYIPEVFT--GVYPP 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 TDTDFINTGNLVKYLNLNHS--EKFTGYPIIDNYSYRGFYOKTHLSYOEYFKVFP 236

QY 303 YAGGGFGLYSGPALLRLYSATSRVHLYPDIDVYGMCMQKLGKGL---VPEKHKGFRTE 359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 YCSGLGLYSDRLVPRYEMGHVKPIKFDVYVIGICLNLLKLVNIHIPEDINLFLYRIH 296

QY 360 ENKNKNICSIDLMVHSRKPQEMIDINSLQSNLKC 397
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 ----LDVCLRRVIAAHGFSSKEIITFW-QVMLRNTIC 329

RESULT 7
US-09-055-097-1
; Sequence 1, Application US/09055097
; Patent No. 5955282
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR

; PRIOR APPLICATION NUMBER: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,097
; FILING DATE: Filed Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cetroni, Michael C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0490 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 378 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: DUODNOT02
; CLONE: 1705085
; US-09-055-097-1

Query Match      13.5%; Score 287; DB 2; Length 378;
Best Local Similarity 28.0%; Pred. No. 1.5e-22;
Matches 84; Conservative 45; Mismatches 117; Indels 54; Gaps 9;

QY 130 LLIDQPKKCA--KKPFLLLAIKSLIPHFARQAIRESMGRETNGVTVVVFLLGKTP 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 LLIPNEAGSGCAPPFLLILVCTAPENLNQRNAIRASWGLREARGLRVOTFLPGE--- 113

QY 187 PEDNHP-----DLSDMLKFESDKHQDILMMNYRDTFENLSLKEVLFLRWVYSTCPDAE 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 114 PNAQHPVWMSGQSDLAS----ESAAQGDILQAAFOOSYRNLTSLTSLGLNMAEKHCPMAR 169

QY 240 FVFKGDDDFVFNTHHLYNLSLSKSKAKDLFGD 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 170 YVUKTDDVVVNVVPELVSELVLRGGRWGWERSTPEQAEQEGGVHSEVPLLYLOR 229

QY 275 VHNAGPHRDKLKYIYIPEVFT--GVYPPYAGGGFGLYSGPALLRLYSATSRVHLXPI 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 VHMVRNPSRTPGGRHVRSEEQPHPTWGPFPYASGTGYVLSASAVQLILKVASRAPLLPL 289

QY 332 DDVYTCMCKLQKLGVP---KHGKFTFDIEKNKNKNICSYDLMV-HSRKPKQEMIDW 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 EDVFGVSARRGGLAPTQCVKLAGATHYPLDR-----CCYKFLTSHRLDPWKMQEA 343

RESULT 8
US-08-207-904-10
; Sequence 10, Application US/08207904
; Patent No. 5477002
; GENERAL INFORMATION:
; APPLICANT: Tuttle, AnnMarie
; APPLICANT: Crossland, Lyle D.
; TITLE OF INVENTION: Anthr-Specific cDNA Sequences, Genomic
; TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
; NUMBER OF SEQUENCES: 21
```

```
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,904
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/908,242
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: CGC 1624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8615
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 372 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-207-904-10

Query Match 6.7%; Score 142; DB 1; Length 372;
Best Local Similarity 20.5%; Pred. No. 8.5e-07;
Matches 72; Conservative 73; Mismatches 138; Indels 68; Gaps 15;

QY 47 KFMKPPST-----PRAYNRQEKLNRMYPILNKNVANQTGELATSPNTSHLSYCEP 98
DB 5 QIWTQPNSEMNNDLVIPAFINHEKLTISRECDPKRLAESNDIMGEVKKTHQATQSL 64
QY 99 DSTVMTAVTDF-----NNLPDRPKDFLLYLRCNYSLLIDQPKKCAKPPFLLLAI 148
DB 65 DKMSSTLENAIARTROTISHNAKEN-----RASNHT-----TPNKAFIGVGI 108
QY 149 KSLIPHARRQAIRESWG-----RETVMGQIVVVRVFLLGKTPEDNHDLSMLKPFES 202
DB 109 NTAFSSRRKRSRLRETNMPKDGKRLKLEKEGIVIR-FVIGHSATRGGVLDRA--DUSED 145
QY 203 DKHODILMNNYRDTFFNLSLKEVLFRLRWVSTSCPDAEFVFVKGGDDDFVFNTHHILNLSL 262
DB 166 AQYKDFLRDHVEGVHELSTRLVFS-KAVSIWDADFYVKVDDVDVHLNLGML---ANTL 221
QY 263 SKSKAKD-LFTGDVHNAGPRDRKK-LKYIYPEVYTG-----VYPPYAGGGGLYSGLPAL 316
DB 222 AKYKSKPRVYIGCM-KSGPVLSQGVRYEYEPYKFEQGNKYFRHATQIYGISRDLA 279
QY 317 LRLYSATSRVHLPIDDVYTGMCQKLGKLVPEKHKGFRTDIEKNNKNIC 367
DB 280 SYISINSGLIHRANEDVSLGSWL--IGL-----EVEHVDERSMC 317

RESULT 9
US-09-464-035A-1
; Sequence 1, Application US/09464035A
; Patent No. 6492152
; GENERAL INFORMATION:
; APPLICANT: Canfield et al.
; TITLE OF INVENTION: Core 1 3-Galactosyl Transferases & Methods of Use
; FILE REFERENCE: 5820.534
; CURRENT APPLICATION NUMBER: US/09/464,035A
; CURRENT FILING DATE: 1999-12-15

; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: WordPerfect 8.0 *Software*
; SEQ ID NO 1
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-464-035A-1

Query Match 4.8%; Score 103; DB 4; Length 363;
Best Local Similarity 20.9%; Pred. No. 0.014;
Matches 67; Conservative 43; Mismatches 102; Indels 108; Gaps 17;

QY 76 RVANOTGELATSPNTSHLSYCEPSTVMTAVTDFNNLPDRFKD-----FLYLRCRNY 129
DB 34 KVDTPNVLHNDPHARHSD--DNGQNHLEGOMNFNADSSQHKDENTDIAENLYQKVRILC 91
QY 130 LLIDOPKKCAKPPFLLLAIKSLIPHARRQAIRESWGRETVMGQIVVVRVFLLGKTPED 189
DB 92 WYMTGPONLEKK-----AKHVKAQWQRN-----KVLFMSS---EE 125
QY 190 NHPDLSMDLKFSQKHQDILMNNYRDTFFNLSLKEVLFRLRWVSTSCPDAEFVFVKGGDDV 248
DB 126 NKDFPAVGLTKEGRDQ--LYW-----KTAKAFQYVHEHYLEDADWFLKAODDT 172
QY 249 FVNTHHILNLSLSKAKDLFIGDVIHNAGPHRDKLKYIPE--VFYTGVPYPPY--- 303
DB 173 YV---ILDNLRLW-----LSKYDPEEPIYFGRRKFPVKQ 204
QY 304 ---AGGGFLYSGPALLRLYSA--TSR-VHLYPIDDDVYTGMCQKLGK----- 345
DB 205 GYMGGAGYVLSKEALKREVDAFKTKCTHSSSIDLALGRMEIMNVNVEAGDSRDTICKE 264
QY 346 -----VPEKHH--KGF--RTF 356
DB 265 TFHPFVPEHHLIKGYLPRTF 284

RESULT 10
US-09-849-562A-1
; Sequence 1, Application US/09849562A
; Patent No. 6545123
; GENERAL INFORMATION:
; APPLICANT: Cummings, Richard D.
; APPLICANT: McEever, Rodger P.
; TITLE OF INVENTION: synthetic glycosulfopeptides and methods of synthesis thereo
; FILE REFERENCE: 5820.594
; CURRENT APPLICATION NUMBER: US/09/849,562A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 09/334,013
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-849-562A-1

Query Match 4.8%; Score 103; DB 4; Length 363;
Best Local Similarity 20.9%; Pred. No. 0.014;
Matches 67; Conservative 43; Mismatches 102; Indels 108; Gaps 17;

QY 76 RVANOTGELATSPNTSHLSYCEPSTVMTAVTDFNNLPDRFKD-----FLYLRCRNY 129
DB 34 KVDTPNVLHNDPHARHSD--DNGQNHLEGOMNFNADSSQHKDENTDIAENLYQKVRILC 91
QY 130 LLIDOPKKCAKPPFLLLAIKSLIPHARRQAIRESWGRETVMGQIVVVRVFLLGKTPED 189
DB 92 WYMTGPONLEKK-----AKHVKAQWQRN-----KVLFMSS---EE 125
QY 190 NHPDLSMDLKFSQKHQDILMNNYRDTFFNLSLKEVLFRLRWVSTSCPDAEFVFVKGGDDV 248
DB 126 NKDFPAVGLTKEGRDQ--LYW-----KTAKAFQYVHEHYLEDADWFLKAODDT 172
```

```

QY 249 FVNTTHILNYLSNLSKSKAKDLFIGDVHNNAGPHRDKKLKYYPE--VFYTGVPYPY--- 303
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 173 YV-----ILDNLRLW-----LSKYDPEEPIYVGRFRFXPVVKQ 204
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 304 ---AGGGGLYSGPALLLYSA--TSR-VHLYPIDDDVYTCMLQKLG----- 345
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 205 GYMGGAGVLSKEALKRFVDAFTDKCTHSSIEDLALGRCHEIMNVVAGDSRDTIGKE 264
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 346 -----VPEKH--KGF--RTF 356
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 265 TFHPFVPEHHLIKGYLPRTF 284
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

RESULT 11
US-09-849-031A-1
; Sequence 1, Application US/09849031A
; Patent No. 6569998
; GENERAL INFORMATION:
; APPLICANT: Cummings, Richard D.
; APPLICANT: McEver, Rodger P.
; TITLE OF INVENTION: Synthetic glycosulfopeptides and methods of synthesis thereof
; FILE REFERENCE: 5820.593
; CURRENT APPLICATION NUMBER: US/09/849,031A
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 09/334,013
; PRIOR FILING DATE: 1993-06-15
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-849-031A-1

Query Match 4.8%; Score 103; DB 4; Length 363;
Best Local Similarity 20.9%; Pred. NO. 0.014;
Matches 67; Conservative 43; Mismatches 102; Indels 108; Gaps 17;

QY 76 RVANQTGELATSPNTSHLSYCEPDSTVMTAVTDFFNLNLPDRFKD-----FLYLRCRNYS 129
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 34 KYDTQPNVLNDPHARHSD--DNGONHLEGOMNFNADSSOHKDENTDIAENLKQVRLC 91
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 130 LLIDQPKCAKPFLLAIAKSLIPHARQAIRSWGRETNGVQTVRVVFLGKTPTD 189
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 92 WYMTGPQNLEK-----AKHVKATWAQRGN-----KVLEMS-----EE 125
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 190 NHPDLSDMLKEESDKHQDILMNYRDTFTFNLSLKEVFLRWVSTS-CPDAEFVFKGDDGV 248
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 126 NKDFPVAVLKTEGRDQ--LVH-----KTIKAFQVHVHEHVEDADFLKADDDT 172
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 249 FVNTTHILNYLSNLSKSKAKDLFIGDVHNNAGPHRDKKLKYIYPE--VFYTGVPYPY--- 303
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 173 YV-----ILDNLRLW-----LSKYDPEEPIYVGRFRFXPVVKQ 204
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 304 ---AGGGGLYSGPALLLYSA--TSR-VHLYPIDDDVYTCMLQKLG----- 345
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 205 GYMGGAGVLSKEALKRFVDAFTDKCTHSSIEDLALGRCHEIMNVVAGDSRDTIGKE 264
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 346 -----VPEKH--KGF--RTF 356
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 265 TFHPFVPEHHLIKGYLPRTF 284
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

RESULT 12
US-09-464-035A-11
; Sequence 11, Application US/09464035A
; Patent No. 6492152
; GENERAL INFORMATION:
; APPLICANT: Canfield et al.
; TITLE OF INVENTION: Core 1 3-Galactosyl Transferases & Methods of Use
; FILE REFERENCE: 5820.534
; CURRENT APPLICATION NUMBER: US/09/464,035A

```

Db 199 KPVYKQYMGAGYVLSKEALRRVDAFKTEKTHSSSIEDLALRCMEIKVEAGDSR 258
QY 346 -----VPEKH--KGF--RTF 356
Db 259 DPTGKETFPFVPEHHLIKGYLPKTF 284

RESULT 14

US-09-464-035A-5
; Sequence 5, Application US/09464035A
; Patent No. 6492152
; GENERAL INFORMATION:
; APPLICANT: Canfield et al.
; TITLE OF INVENTION: Core 1 3-Galactosyl Transferases & Methods of Use
; FILE REFERENCE: 5820.534
; CURRENT APPLICATION NUMBER: US/09/464.035A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: WordPerfect 8.0 *Software*
; SEQ ID NO 5
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-464-035A-5

Query Match 4.6%; Score 97; DB 4; Length 363;
Best Local Similarity 20.4%; Pred. No. 0.064;
Matches 55; Conservative 41; Mismatches 73; Indels 100; Gaps 15;
QY 121 LYLRCRNYSLLDQPKKCAKPFLLLAIKSLIPHFARROAIRSNGKRETNVGNQTVRVYF 180
Db 83 LYQVKVILCWVWTSFONLEK-----AKVVKATVQRN-----KVL 119
QY 181 LLGKTPPDNHPDLSDMLKPSDKHODILMNYRDFTFNLKLVFLRWVTS-CPDAE 239
Db 120 FMSS---EENQDFTVGLTKEGREQ--LYW-----KTIKAFQVVDHYLEDAD 163
QY 240 FVFGKDDVDVFNTHHILNYNSLSKSKAKDLFIGDVHINAGPHRDKKLKYYIPE--VEYT 297
Db 164 WFMKADDTYV-----IVDNLWL-----LSKNPEQDIYEG 195
QY 298 GVYPPY-----AGGGGLYSGPALLRLYSA--TSR-VHLYPIDDYVTGMLQKGLG--- 345
Db 196 RRFEPYKQYMGAGYVLSKEALRRVDAFKTEKTHSSSIEDLALRCMEIKVEAG 255
QY 346 -----VPEKH--KGF--RTF 356
Db 259 DPTGKETFPFVPEHHLIKGYLPKTF 284

RESULT 15

US-08-756-317-9
; Sequence 9, Application US/08756317
; Patent No. 5849894
; GENERAL INFORMATION:
; APPLICANT: Clemente, Thomas E.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Mitsky, Timothy A.
; APPLICANT: Stark, David M.
; TITLE OF INVENTION: Improved Rhodospirillum Rubrum
; TITLE OF INVENTION: Poly-B-Hydroxyalkanoate Synthesis
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756.317
; FILING DATE: 25-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,693
; FILING DATE: 29-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE/DOCKET NUMBER: MOBT:008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 787-1400
; TELEFAX: (713) 787-1440
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 624 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-756-317-9

Query Match 4.1%; Score 88; DB 2; Length 624;
Best Local Similarity 20.3%; Pred. No. 1.4; Indels 88; Gaps 12;
Matches 58; Conservative 41; Mismatches 41; Indels 88; Gaps 12;
QY 49 WKPFSTPRAYNNREQKLNRYNPIILNRVANQTGELATSPNTSHLSYCEPDSTV----- 102
Db 322 WKNPDP--SYGDTGMDYVSAYLEVMRDVL---DLTDQKLNNAVGYCIAGTTLALIPVV 375
QY 103 -----MTAVTDFNLLPDRFKDFLLYLRCRNYSLLDQPKKCAKPFLLLA 148
Db 376 LKQGRDVRVNAATFTALTDFADQ---GEFTAYLOEDFVSGIEEAARTG-----ILGA 426
QY 149 KSLIPHFARROAIRSNGKRETNVGNQTVRVFLGKTPPDNHPDLSDMLKPSDKHODI 208
Db 427 QLMTRTSEFLRANDLVWG-----PAIRSYMLGETPPA-----FDL 461
QY 209 LMWNYRDTFFNLKLVFLRWVTSICPDAEFFVKGDG-----DYFV-----NT 252
Db 462 LFVWNGDGT---NLPGRMAVEYLRGLCQQRNFVKEGFDLMGHLHVGDDVTVPJCAIACET 517
QY 253 RHLLNLYNS-----LSKSKAKDLFIGDVHINAG---PHRDKKLKYY 290
Db 518 DRIAPWKDSWRGIAQMGSRQKIFILSESGHIAIGIVNPPSKKXYGHY 563

Search completed: October 20, 2003, 23:05:18
Job time : 13.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 23:04:10 ; Search time 30 Seconds
(without alignments)
2169.170 Million cell updates/sec

Title: US-09-804-357B-2
Perfect score: 2129
Sequence: 1 MSVGRRRVKLLGILMANVF.....RKQEMIDWSQLSPNLKC 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 609560 seqs, 163917102 residues

Total number of hits satisfying chosen parameters: 609560

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PTI_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2129	100.0	397	9	US-09-804-357-2
2	2129	100.0	397	10	US-09-804-006-2
3	2113	99.2	397	14	US-10-109-563-2
4	1979	93.0	365	9	US-09-804-357-4
5	1979	93.0	365	10	US-09-804-006-4
6	1862	87.5	397	9	US-09-804-357-14
7	1862	87.5	397	10	US-09-972-912-2
8	1862	87.5	397	10	US-09-804-006-14
9	1862	87.5	397	14	US-10-109-563-4
10	1725	81.0	365	9	US-09-804-357-15
11	1725	81.0	365	10	US-09-804-006-15
12	656	30.8	352	9	US-09-739-451-4
13	656	30.8	353	9	US-09-739-451-12
14	654.5	30.7	353	15	US-10-097-065-137
15	654.5	30.7	401	12	US-10-288-252-10

16	647.5	30.4	378	12	US-10-237-496-36	Sequence 36, Appl
17	647.5	30.4	378	12	US-10-242-074-36	Sequence 36, Appl
18	647.5	30.4	378	12	US-10-242-505-36	Sequence 36, Appl
19	647.5	30.4	378	12	US-10-242-574-36	Sequence 36, Appl
20	647.5	30.4	378	12	US-10-243-261-36	Sequence 36, Appl
21	647.5	30.4	378	12	US-10-243-282-36	Sequence 36, Appl
22	647.5	30.4	378	12	US-10-243-402-36	Sequence 36, Appl
23	647.5	30.4	378	12	US-10-243-431-36	Sequence 36, Appl
24	647.5	30.4	378	12	US-10-245-164-36	Sequence 36, Appl
25	647.5	30.4	378	12	US-10-244-972-36	Sequence 36, Appl
26	647.5	30.4	378	12	US-10-197-942-36	Sequence 36, Appl
27	647.5	30.4	378	12	US-10-238-196-36	Sequence 36, Appl
28	647.5	30.4	378	12	US-10-245-013-36	Sequence 36, Appl
29	647.5	30.4	378	15	US-10-245-103-36	Sequence 36, Appl
30	647.5	30.4	378	15	US-10-245-107-36	Sequence 36, Appl
31	647.5	30.4	378	15	US-10-245-143-36	Sequence 36, Appl
32	647.5	30.4	378	15	US-10-245-771-36	Sequence 36, Appl
33	647.5	30.4	378	15	US-10-245-851-36	Sequence 36, Appl
34	647.5	30.4	378	15	US-10-245-883-36	Sequence 36, Appl
35	647.5	30.4	378	15	US-10-237-535-36	Sequence 36, Appl
36	647.5	30.4	378	15	US-10-238-183-36	Sequence 36, Appl
37	647.5	30.4	378	15	US-10-238-283-36	Sequence 36, Appl
38	647.5	30.4	378	15	US-10-238-370-36	Sequence 36, Appl
39	647.5	30.4	378	15	US-10-245-055-36	Sequence 36, Appl
40	647.5	30.4	378	15	US-10-245-147-36	Sequence 36, Appl
41	647.5	30.4	378	15	US-10-245-730-36	Sequence 36, Appl
42	647.5	30.4	378	15	US-10-245-739-36	Sequence 36, Appl
43	647.5	30.4	378	15	US-10-246-210-36	Sequence 36, Appl
44	647.5	30.4	378	15	US-10-239-196-36	Sequence 36, Appl
45	647.5	30.4	378	15	US-10-243-024-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-09-804-357-2
; Sequence 2, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-804-357-2

Query Match: 100.0%; Score 2129; DB 9; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.6e-208;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1 MSVGRRRVKLLGILMANVFYILVEYKNSODKNGKGVIIIPKEKFKWKPSTPRAYWN 60
Db 1 MSVGRRRVKLLGILMANVFYILVEYKNSODKNGKGVIIIPKEKFKWKPSTPRAYWN 60
Qy 61 REQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRPKDEL 120
Db 61 REQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRPKDEL 120


```
QY 121 LYLRCRNYSLLDOPKCKAKKFFLLAIKSLIPHFAKQA:RESWGRETNGNOTVVRVF 180
Db 121 LYLRCRNYSLLDOPKCKAKKFFLLAIKSLIPHFAKQA:RESWGRETNGNOTVVRVF 180
QY 181 LLGKTPEDNHPDLSMDLKFESDKHODILMWNRYRTFFNLSLKEVLFLLRWVSTSCPDADF 240
Db 181 LLGKTPEDNHPDLSMDLKFESDKHODILMWNRYRTFFNLSLKEVLFLLRWVSTSCPDADF 240
QY 241 VFKGDDVFNTHHILNYLNSLSKSKAKDLFGDVIHNAHPHRDKKLYIPEVFTGYV 300
Db 241 VFKGDDVFNTHHILNYLNSLSKSKAKDLFGDVIHNAHPHRDKKLYIPEVFTGYV 300
QY 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDVYTGMCLOKLGVLPEKHKGFRFTDIEE 360
Db 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDVYTGMCLOKLGVLPEKHKGFRFTDIEE 360
QY 361 KKKKNICSYIDLMLVHSRKPQEMIDWISOLQSPNLKC 397
Db 361 KKKKNICSYIDLMLVHSRKPQEMIDWISOLQSPNLKC 397

RESULT 2
US-09-804-006-2
; Sequence 2, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/126001
; CURRENT APPLICATION NUMBER: US/09/804,006
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/292,228
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-804-006-2

Query Match 100.0%; Score 2129; DB 10; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.6e-208;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVGRRRVKLLGILMMANFYIYLIVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYN 60
Db 1 MSVGRRRVKLLGILMMANFYIYLIVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYN 60
QY 61 REQEKLRWYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDFL 120
Db 61 REQEKLRWYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDFL 120
QY 121 LYLRCRNYSLLDOPKCKAKKFFLLAIKSLIPHFAKQA:RESWGRETNGNOTVVRVF 180
Db 121 LYLRCRNYSLLDOPKCKAKKFFLLAIKSLIPHFAKQA:RESWGRETNGNOTVVRVF 180
QY 181 LLGKTPEDNHPDLSMDLKFESDKHODILMWNRYRTFFNLSLKEVLFLLRWVSTSCPDADF 240
Db 181 LLGKTPEDNHPDLSMDLKFESDKHODILMWNRYRTFFNLSLKEVLFLLRWVSTSCPDADF 240
QY 241 VFKGDDVFNTHHILNYLNSLSKSKAKDLFGDVIHNAHPHRDKKLYIPEVFTGYV 300
Db 241 VFKGDDVFNTHHILNYLNSLSKSKAKDLFGDVIHNAHPHRDKKLYIPEVFTGYV 300
QY 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDVYTGMCLOKLGVLPEKHKGFRFTDIEE 360
Db 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDVYTGMCLOKLGVLPEKHKGFRFTDIEE 360
```

```
Db 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDVYTGMCLOKLGVLPEKHKGFRFTDIEE 360
QY 361 KKKKNICSYIDLMLVHSRKPQEMIDWISOLQSPNLKC 397
Db 361 KKKKNICSYIDLMLVHSRKPQEMIDWISOLQSPNLKC 397

RESULT 3
US-10-109-563-2
; Sequence 2, Application US/10109563
; Publication No. US20020170075A1
; GENERAL INFORMATION:
; APPLICANT: Leviten, Michael W.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE COMPRISING
; FILE REFERENCE: BETA-1,3-N-ACETYLLACTOSAMINE DISRUPTIONS
; CURRENT APPLICATION NUMBER: US/10/109,563
; CURRENT FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: US 60/280,706
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-109-563-2

Query Match 99.2%; Score 2113; DB 14; Length 397;
Best Local Similarity 99.2%; Pred. No. 6.9e-207;
Matches 394; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSVGRRRVKLLGILMMANFYIYLIVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYN 60
Db 1 MSVGRRRVKLLGILMMANFYIYLIVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYN 60
QY 61 REQEKLRWYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDFL 120
Db 61 REQEKLRWYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDFL 120
QY 121 LYLRCRNYSLLDOPKCKAKKFFLLAIKSLIPHFAKQA:RESWGRETNGNOTVVRVF 180
Db 121 LYLRCRNYSLLDOPKCKAKKFFLLAIKSLIPHFAKQA:RESWGRETNGNOTVVRVF 180
QY 181 LLGKTPEDNHPDLSMDLKFESDKHODILMWNRYRTFFNLSLKEVLFLLRWVSTSCPDADF 240
Db 181 LLGKTPEDNHPDLSMDLKFESDKHODILMWNRYRTFFNLSLKEVLFLLRWVSTSCPDADF 240
QY 241 VFKGDDVFNTHHILNYLNSLSKSKAKDLFGDVIHNAHPHRDKKLYIPEVFTGYV 300
Db 241 VFKGDDVFNTHHILNYLNSLSKSKAKDLFGDVIHNAHPHRDKKLYIPEVFTGYV 300
QY 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDVYTGMCLOKLGVLPEKHKGFRFTDIEE 360
Db 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDVYTGMCLOKLGVLPEKHKGFRFTDIEE 360
QY 361 KKKKNICSYIDLMLVHSRKPQEMIDWISOLQSPNLKC 397
Db 361 KKKKNICSYIDLMLVHSRKPQEMIDWISOLQSPNLKC 397

RESULT 4
US-09-804-357-4
; Sequence 4, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
```

; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-804-357-4

Query Match 93.0%; Score 1979; DB 9; Length 365;
Best Local Similarity 100.0%; Pred. No. 2.9e-193;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 QDKNGKGGVVIIPKPKFKPKPPSTPRAYWNRQEKLNRYNPFILNRVANOTGELATSPNTSH 92
DB 1 QDKNGKGGVVIIPKPKFKPKPPSTPRAYWNRQEKLNRYNPFILNRVANOTGELATSPNTSH 60

QY 93 LSYCEPDSVTMTAVTDFNNLPDRFKDFLLYLRCRNYSLLDOPKKCAKPFLLAIAKSLI 152
DB 61 LSYCEPDSVTMTAVTDFNNLPDRFKDFLLYLRCRNYSLLDOPKKCAKPFLLAIAKSLI 120

QY 153 PHFARRQAIRESWGRETNGVQTVVRVFLGKTPPEDNHPDLSMDLKFESDKHODILMWN 212
DB 121 PHFARRQAIRESWGRETNGVQTVVRVFLGKTPPEDNHPDLSMDLKFESDKHODILMWN 180

QY 213 YRDTFFNLSLKEVLFLRWVSTSCPDAEVFKGDDVFNTHHILNLYNSLSKSKAKDLFI 272
DB 181 YRDTFFNLSLKEVLFLRWVSTSCPDAEVFKGDDVFNTHHILNLYNSLSKSKAKDLFI 240

QY 273 GDVIHNAHPHRDKKLKYIPEVFTGYPPYAGGGFLYSGPALLRLLYSATSRVHLYPID 332
DB 241 GDVIHNAHPHRDKKLKYIPEVFTGYPPYAGGGFLYSGPALLRLLYSATSRVHLYPID 300

QY 333 DVTGMCLOKLGVLPEKHKGFRFTDIEKNKKNICSVIDMLVHSRKPQEMIDIWSQLOS 392
DB 301 DVTGMCLOKLGVLPEKHKGFRFTDIEKNKKNICSVIDMLVHSRKPQEMIDIWSQLOS 360

QY 393 PNLKC 397
DB 361 PNLKC 365

RESULT 5
US-09-804-006-4
; Sequence 4, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/126001
; CURRENT APPLICATION NUMBER: US/09/804,006
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/292,228
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-804-006-4

Query Match 93.0%; Score 1979; DB 10; Length 365;
Best Local Similarity 100.0%; Pred. No. 2.9e-193;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 QDKNGKGGVVIIPKPKFKPKPPSTPRAYWNRQEKLNRYNPFILNRVANOTGELATSPNTSH 92
DB 1 QDKNGKGGVVIIPKPKFKPKPPSTPRAYWNRQEKLNRYNPFILNRVANOTGELATSPNTSH 60

QY 93 LSYCEPDSVTMTAVTDFNNLPDRFKDFLLYLRCRNYSLLDOPKKCAKPFLLAIAKSLI 152
DB 61 LSYCEPDSVTMTAVTDFNNLPDRFKDFLLYLRCRNYSLLDOPKKCAKPFLLAIAKSLI 120

QY 153 PHFARRQAIRESWGRETNGVQTVVRVFLGKTPPEDNHPDLSMDLKFESDKHODILMWN 212
DB 121 PHFARRQAIRESWGRETNGVQTVVRVFLGKTPPEDNHPDLSMDLKFESDKHODILMWN 180

QY 213 YRDTFFNLSLKEVLFLRWVSTSCPDAEVFKGDDVFNTHHILNLYNSLSKSKAKDLFI 272
DB 181 YRDTFFNLSLKEVLFLRWVSTSCPDAEVFKGDDVFNTHHILNLYNSLSKSKAKDLFI 240

QY 273 GDVIHNAHPHRDKKLKYIPEVFTGYPPYAGGGFLYSGPALLRLLYSATSRVHLYPID 332
DB 241 GDVIHNAHPHRDKKLKYIPEVFTGYPPYAGGGFLYSGPALLRLLYSATSRVHLYPID 300

QY 333 DVTGMCLOKLGVLPEKHKGFRFTDIEKNKKNICSVIDMLVHSRKPQEMIDIWSQLOS 392
DB 301 DVTGMCLOKLGVLPEKHKGFRFTDIEKNKKNICSVIDMLVHSRKPQEMIDIWSQLOS 360

QY 393 PNLKC 397
DB 361 PNLKC 365

RESULT 6
US-09-804-357-14
; Sequence 14, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-804-357-14

Query Match 87.5%; Score 1862; DB 9; Length 397;
Best Local Similarity 86.9%; Pred. No. 2.9e-181;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;

QY 1 MSVGRRRVKKLIGILMMANVFYILVEVSKNSQDKNGKGGVVIIPKFKWKPSTPRAYWN 60
DB 1 MSVGRRIKKLIGILMMANVFYIFMEVSKSSQSKNGKEVVIIPKFKWISTPPEAYWN 60

QY 61 REQEKLNRYNPFILNRVANOTGELATSPNTSHLSYCEPDSVTMTAVTDFNNLPDRFKDL 120
DB 61 REQEKLNRYNPFILNLTNQTGEAGRUSNLSHLSYCEPDSVTMTAVTDFNNLPDRFKDL 120

QY 121 LYLRCRNYSLLDOPKKCAKPFLLAIAKSLIPHARQAIRESWGRETNGVQTVVRV 180

```
Db 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARRQAIRESWGQESNAGNQTWVRVF 180
Qy 181 LLGKTPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFRLRWVSTSCPDADF 240
Db 181 LLGKTPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFRLRWVSTSCPDADF 240
Qy 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFIDGVIHNNAGPHRDKKLYIPEVFTGYV 300
Db 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFIDGVIHNNAGPHRDKKLYIPEVFTGYV 300
Qy 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDDVTGMCLOKGLVPEKHKGFRTFIDE 360
Db 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDDVTGMCLOKGLVPEKHKGFRTFIDE 360
Qy 361 KNKKNICSYIDLMLVHSRKPQEMIDWISQSPNLKC 397
Db 361 KNKKNICSYVDLMLVHSRKPQEMIDWISQSAHLKC 397

RESULT 7
US-09-972-912-2
: Sequence 2, Application US/09972912
: Patent No. US20020110867A1
: GENERAL INFORMATION:
: APPLICANT: SOPPET, DANIEL R.
: RUBEN, STEVEN M.
: TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
: STREET: 1100 NEW YORK AVENUE, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: US
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA: US/09/972,912
: APPLICATION NUMBER: US/09/972,912
: FILING DATE: 10-Oct-2001
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/049,022
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: STEFFE, ERIC K.
: REGISTRATION NUMBER: 36,668
: REFERENCE/DOCKET NUMBER: 1488.0620001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 397 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-972-912-2

Query Match 87.5%; Score 1862; DB 10; Length 397;
Best Local Similarity 86.9%; Pred. No. 2,9e-181;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;

Qy 1 MSVGRRRVKKLLGILIMANVFYIIVEVSKSSQDKNGKGVIIIPKEKFWKPPSTPRAYN 60
Db 1 MSVGRRRVKKLLGILIMANVFYIIVEVSKSSQDKNGKGVIIIPKEKFWKPPSTPRAYN 60
Qy 61 REQEKLNRYNPIILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVTFNNLPDRFKDFL 120
Db 61 REQEKLNRYNPIILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVTFNNLPDRFKDFL 120
Qy 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARRQAIRESWGQESNAGNQTWVRVF 180
Db 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARRQAIRESWGQESNAGNQTWVRVF 180
Qy 181 LLGKTPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFRLRWVSTSCPDADF 240
Db 181 LLGKTPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFRLRWVSTSCPDADF 240
Qy 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFIDGVIHNNAGPHRDKKLYIPEVFTGYV 300
Db 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFIDGVIHNNAGPHRDKKLYIPEVFTGYV 300
```

```
Db 61 REQEKLNRYNPIILSMITNOTGEAGRLSNISHLNYPEDLRVTSVVTGNNLPDRFKDFL 120
Qy 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARRQAIRESWGQESNAGNQTWVRVF 180
Db 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARRQAIRESWGQESNAGNQTWVRVF 180
Qy 181 LLGKTPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFRLRWVSTSCPDADF 240
Db 181 LLGKTPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFRLRWVSTSCPDADF 240
Qy 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFIDGVIHNNAGPHRDKKLYIPEVFTGYV 300
Db 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFIDGVIHNNAGPHRDKKLYIPEVFTGYV 300
Qy 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDDVTGMCLOKGLVPEKHKGFRTFIDE 360
Db 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDDVTGMCLOKGLVPEKHKGFRTFIDE 360
Qy 361 KNKKNICSYIDLMLVHSRKPQEMIDWISQSPNLKC 397
Db 361 KNKKNICSYVDLMLVHSRKPQEMIDWISQSAHLKC 397

RESULT 8
US-09-804-006-14
: Sequence 14, Application US/09804006
: Patent No. US20020119517A1
: GENERAL INFORMATION:
: APPLICANT: White, David
: APPLICANT: Zhou, Jianghong
: TITLE OF INVENTION: LEPTIN INDUCED GENES
: FILE REFERENCE: 07334/126001
: CURRENT APPLICATION NUMBER: US/09/804,006
: CURRENT FILING DATE: 2001-03-12
: PRIOR APPLICATION NUMBER: US 09/292,228
: PRIOR FILING DATE: 1999-04-15
: PRIOR APPLICATION NUMBER: US 60/108,379
: PRIOR FILING DATE: 1998-10-29
: PRIOR APPLICATION NUMBER: US 09/150,857
: PRIOR FILING DATE: 1998-09-10
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 14
: LENGTH: 397
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-804-006-14

Query Match 87.5%; Score 1862; DB 10; Length 397;
Best Local Similarity 86.9%; Pred. No. 2,9e-181;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;

Qy 1 MSVGRRRVKKLLGILIMANVFYIIVEVSKSSQDKNGKGVIIIPKEKFWKPPSTPRAYN 60
Db 1 MSVGRRRVKKLLGILIMANVFYIIVEVSKSSQDKNGKGVIIIPKEKFWKPPSTPRAYN 60
Qy 61 REQEKLNRYNPIILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVTFNNLPDRFKDFL 120
Db 61 REQEKLNRYNPIILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVTFNNLPDRFKDFL 120
Qy 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARRQAIRESWGQESNAGNQTWVRVF 180
Db 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARRQAIRESWGQESNAGNQTWVRVF 180
Qy 181 LLGKTPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFRLRWVSTSCPDADF 240
Db 181 LLGKTPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFRLRWVSTSCPDADF 240
Qy 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFIDGVIHNNAGPHRDKKLYIPEVFTGYV 300
Db 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFIDGVIHNNAGPHRDKKLYIPEVFTGYV 300
```

```

: TITLE OF INVENTION: LEPTIN INDUCED GENES
: FILE REFERENCE: 07334/109001
: CURRENT APPLICATION NUMBER: US/09/804,357
: CURRENT FILING DATE: 2001-03-12
: PRIOR APPLICATION NUMBER: US 09/195,896
: PRIOR FILING DATE: 1998-11-19
: PRIOR APPLICATION NUMBER: US 60/108,379
: PRIOR FILING DATE: 1998-10-29
: PRIOR APPLICATION NUMBER: US 09/150,857
: PRIOR FILING DATE: 1998-09-10
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 15
: LENGTH: 365
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-804-357-15

Query Match      81.0%  Score 1725;  DB 9;  Length 365;
Best Local Similarity 86.8%;  Pred. No. 2.4e-167;
Matches 317;  Conservative 16;  Mismatches 32;  Indels 0;  Gaps 0;

Qy      33  QDKNGKGVIIPEKFKWKPSTPRAYNREQEKLARWYNPILNRVANOTGELATSPNTSH 92
      1:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1  QEKNGKGVIIPEKFKWKISTTPEAYNREQEKLARQYNPILSMLTNOTGAGRLSNISH 60

Qy      93  LSYCEPSTVMTAVTDENNLPDRFKDFLLYLRCRNYSLIIDQPKCAKKPFLLLAIKSLI 152
      1:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      61  LNYCEPDLRVTVSVVTGPNFLPDRFKDFLLYLRCRNYSLIIDQPKCAKKPFLLLAIKSLT 120

Qy      153  PHEAFROAIRRESWGRTFNVCNQTWRVFLLGKTPPDNDHPDLSDMLKPFESKHODILMWN 212

```

```

Db      121 PHFAKRAQRESWGQESNAGNQTQVAVFLGQTTPEDNHPDLSOMLKESEKHQDILWNN 180
QY      213 YRDTFFNLSLKEVLFRLRWVSTSCPDAEFVKGDGDDVFVNTHHILNYLNSLSKSAKOLFI 272
Db      181 YRDTFFNLSLKEVLFRLRWVSTSCPDTEFVKGDGDDVFVNTHHILNYLNSLSKTAKOLFI 240
QY      273 GDVIHNAQPHRDKKLYIYIPEVYFTGVYPPYAGGGGFLYSGPALLRLYISATSRVHLYPID 332
Db      241 GDVIHNAQPHRDKKLYIYIPEVYSGLYPPYAGGGGFLYSGHLALRLYHITDQVHLYPID 300
QY      333 DVTYGMCLQKLGVLPEKHKGFRTDTEENKKNKICSYIDMLVHSRKPQEMIDIWSOLOQ 392
Db      301 DVTYGMCLQKLGVLPEKHKGFRTDTEENKKNKICSYIDMLVHSRKPQEMIDIWSOLOQ 360
QY      393 PNLC 397
Db      361 AHLK 365

```

```

; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 365
;

```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-804-006-15

Query Match      81.0%; Score 1725; DB 10; Length 365;
Best Local Similarity 86.8%; Pred. No. 2.4e-167;
Matches 317; Conservative 16; Mismatches 32; Indels 0; Gaps 0;

Qy 33 QDRNGKGVIIPKEKWKPPSTPRAYWNRQEKLNRYNPIILNRVANOTGELATSPNTSH 92
Db 1 QEKNGKEVILIPKEKWKISTPPEAYWNRQEKLNRYNPIILSMLTNQTAGRLSNISH 60
Qy 93 LSYCEPDSTVMTAVTDFNNLPDRFKDPLLYLRCRNTSLLDOPKKCAKPFLLLAIKSLI 152
Db 61 LNYCEPDLRTSVTVGNNLPDRFKDPLLYLRCRNTSLLDOPKKCAKPFLLLAIKSLT 120
Qy 153 PHFARQAIRESWGRETNGVQTVVRFVLLGKTPPEDNHPDLSMDLKESDKHODILMWN 212
Db 121 PHFARQAIRESWGQESNAGNQTVVRFVLLGQTPPEDNHPDLSMDLKESDKHODILMWN 180
Qy 213 YRDTFFNLSLKEVLFLRWVSTSCPDAEFVKGDGDDVFVNTHHILNYLSLSKSKAKDLFI 272
Db 181 YRDTFFNLSLKEVLFLRWVSTSCPDTEFVKGDGDDVFVNTHHILNYLSLSKSKAKDLFI 240
Qy 273 GDVIHAGPHRDKKLYIPEVYTGVPYAGGGFLYSGPALLRLYSATSRVHLYPID 332
Db 241 GDVIHAGPHRDKKLYIPEVYSGLYGPPYAGGGFLYSGHLALRLYHITQVHLYPID 300
Qy 333 DVTGMCLOKLGVLPEKHGKFRFTDIEKNKKNICSYIDLMLVHSRKPQEMIDINWSQSPN 392
Db 301 DVTGMCLOKLGVLPEKHGKFRFTDIEKNKKNICSYIDLMLVHSRKPQEMIDINWSQSPN 360
Qy 393 PNLC 397
Db 361 AHLK 365

RESULT 12
US-09-739-451-4
; Sequence 4, Application US/09739451
; Patent No. US20010024813A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Dendriac and Brainiac-3
; FILE REFERENCE: PF464
; CURRENT APPLICATION NUMBER: US/09/739,451
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/213,364
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/077,687
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/108,928
; PRIOR FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-451-4

Query Match      30.8%; Score 656; DB 9; Length 353;
Best Local Similarity 40.2%; Pred. No. 3e-58;
Matches 146; Conservative 53; Mismatches 122; Indels 42; Gaps 8;

Qy 40 GVIIPEKFKWKPPSTPRAYWNRQEKLNRYNPIILNRVANOTGELATSPNTSHLSYCEPD 99
Db 19 GCLLFRLKAAPETPR-----PTSLSGAPPTPRHSRCPPN 54
Qy 100 STVMTAVTDFNNLPDRFKDPLLYLRCRNTSLLDOPKKCAKPFLLLAIKSLIPHFARQ 159
Db 55 HTVSSASL---SLPSRHLFLTYRHCRNFSILL-EPGSGCKDTFLLLAIKSQPGHVERA 110
Qy 160 AIRESWGRETNGVQTVVRFVLL---CKTPPEDNHPDLSMDLKESDKHODILMWNRYDT 216
Db 111 AIRSTWGRWGDLGPGALKVFLGCVAGSAPP-----AQLLAYESREFDDILQWDFTE 163
Qy 217 FFNLSLKEVLFLRWVSTSCPDAEFVKGDGDDVFVNTHHILNYLSLSKSKAKDLFI 276
Db 164 FFNLTJLKLHLQWVAACPAHFMLKGGDDVFVHPVNVLEFLDGM--DPAQDLLVGDVI 221
Qy 277 HNAGPHRDKKLYIPEVYTGVPYAGGGFLYSGPALLRLYSATSRVHLYPID 335
Db 222 QOALPNENTKVKYFIPPSMYRATHYPPYAGGGVYMSRATVRRLQAIMEDAELFIDDFV 281
Qy 336 TGMCLQKLGVLPEKHGKFRFTDIEEK-NKKNICSYIDLMLVHSRKPQEMIDINWSQSPN 394
Db 282 VGMCLRLGLSPMHAGFKFTGIRRLDPLDPLCLYRGLLLVHRLSPLEMTWMLVYDDEG 341
Qy 395 LK 397
Db 342 LK 344

RESULT 13
US-09-739-451-12
; Sequence 12, Application US/09739451
; Patent No. US20010024813A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Dendriac and Brainiac-3
; FILE REFERENCE: PF464
; CURRENT APPLICATION NUMBER: US/09/739,451
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/213,364
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/077,687
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/108,928
; PRIOR FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-451-12

Query Match      30.8%; Score 656; DB 9; Length 353;
Best Local Similarity 40.2%; Pred. No. 3e-58;
Matches 146; Conservative 53; Mismatches 122; Indels 42; Gaps 8;

Qy 40 GVIIPEKFKWKPPSTPRAYWNRQEKLNRYNPIILNRVANOTGELATSPNTSHLSYCEPD 99
Db 19 GCLLFRLKAAPETPR-----PTSLSGAPPTPRHSRCPPN 54
Qy 100 STVMTAVTDFNNLPDRFKDPLLYLRCRNTSLLDOPKKCAKPFLLLAIKSLIPHFARQ 159
Db 55 HTVSSASL---SLPSRHLFLTYRHCRNFSILL-EPGSGCKDTFLLLAIKSQPGHVERA 110
Qy 160 AIRESWGRETNGVQTVVRFVLL---CKTPPEDNHPDLSMDLKESDKHODILMWNRYDT 216
Db 111 AIRSTWGRWGDLGPGALKVFLGCVAGSAPP-----AQLLAYESREFDDILQWDFTE 163
Qy 217 FFNLSLKEVLFLRWVSTSCPDAEFVKGDGDDVFVNTHHILNYLSLSKSKAKDLFI 276
Db 164 FFNLTJLKLHLQWVAACPAHFMLKGGDDVFVHPVNVLEFLDGM--DPAQDLLVGDVI 221
Qy 277 HNAGPHRDKKLYIPEVYTGVPYAGGGFLYSGPALLRLYSATSRVHLYPID 335
Db 222 QOALPNENTKVKYFIPPSMYRATHYPPYAGGGVYMSRATVRRLQAIMEDAELFIDDFV 281
Qy 336 TGMCLQKLGVLPEKHGKFRFTDIEEK-NKKNICSYIDLMLVHSRKPQEMIDINWSQSPN 394
Db 282 VGMCLRLGLSPMHAGFKFTGIRRLDPLDPLCLYRGLLLVHRLSPLEMTWMLVYDDEG 341
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 13:13:22 ; Search time 4471.3 Seconds
(without alignments)
10942.656 Million cell: updates/sec

Title: US-09-804-357B-1

Perfect score: 1196

Sequence: 1 agatgagtgaggcgctga.....gtcctaatttaaatgctga il96

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.:

1: gb_ba.:

2: gb_htg.:

3: gb_in.:

4: gb_ov.:

5: gb_ov.:

6: gb_pat.:

7: gb_ph.:

8: gb_pl.:

9: gb_pr.:

10: gb_ro.:

11: gb_sts.:

12: gb_sy.:

13: gb_un.:

14: gb_vi.:

15: em_ba.:

16: em_fun.:

17: em_hum.:

18: em_in.:

19: em_mu.:

20: em_mu.:

21: em_or.:

22: em_ov.:

23: em_pat.:

24: em_ph.:

25: em_pl.:

26: em_ro.:

27: em_sts.:

28: em_un.:

29: em_vi.:

30: em_htg_hum.:

31: em_htg_inv.:

32: em_htg_other.:

33: em_htg_mus.:

34: em_htg_pln.:

35: em_htg_rod.:

36: em_htg_mam.:

37: em_htg_vrt.:

38: em_sy.:

39: em_htgo_hum.:

40: em_htgo_mus.:

41: em_htgo_other.:

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1196	100.0	2468	10	BC009075	BC009075 Mus muscu
2	1196	100.0	2672	10	AY043479	AY043479 Mus muscu
3	1196	100.0	210863	10	AL772364	AL772364 Mouse DNA
4	1192.8	99.7	1260	10	AF092050	AF092050 Mus muscu
5	1064.4	89.0	178009	2	AC119549	AC119549 Rattus no
6	1056.4	88.3	318862	2	AC109547	AC109547 Rattus no
7	906.4	75.8	1831	9	AF092051	AF092051 Homo sapi
8	906.4	75.8	1912	6	BD093191	BD093191 Useful po
9	906.4	75.8	1912	9	AB049584	AB049584 Homo sapi
10	906.4	75.8	2454	9	BC030579	BC030579 Homo sapi
11	906.4	75.8	2714	9	AF288208	AF288208 Homo sapi
12	906.4	75.8	2811	9	BC047933	BC047933 Homo sapi
13	906.4	75.8	99088	9	AC093401	AC093401 Homo sapi
14	893	74.7	1197	9	HSA6077	AJ006077 Homo sapi
15	885.4	74.0	2516	9	AF288209	AF288209 Homo sapi
16	848.6	71.0	2442	6	BD156964	BD156964 Primer fo
17	848.6	71.0	2442	9	AK002009	AK002009 Homo sapi
18	724.8	60.6	126480	2	AC040954	AC040954 Mus muscu
19	713.8	59.7	172509	2	AC024716	AC024716 Homo sapi
20	713.8	59.7	200191	9	AL390027	AL390027 Human DNA
21	713.8	59.7	288669	2	AL390297	AL390297 Homo sapi
22	521.8	43.6	832	6	BD149792	BD149792 Primer fo
23	415.4	34.7	170370	2	BX469915	BX469915 Danio rer
24	415	34.7	2162	5	AF321831	AF321831 Danio rer
25	375	31.4	1682	5	AF321828	AF321828 Danio rer
26	370	30.9	556	6	BD071997	BD071997 Secreted
27	362.6	30.3	197598	2	AC144887	AC144887 Bos tauru
28	203.8	17.0	1245	6	AX513603	AX513603 Sequence
29	186.2	15.6	2260	5	AF321827	AF321827 Danio rer
30	185.6	15.5	1564	5	AF321829	AF321829 Danio rer
31	178	14.9	209192	2	AC102609	AC102609 Mus muscu
32	174.8	14.6	1502	10	AF502429	AF502429 Mus muscu
33	174.8	14.6	2367	10	BC031187	BC031187 Mus muscu
34	173.2	14.5	230799	2	AC097668	AC097668 Rattus no
35	172.6	14.4	1434	9	AF502430	AF502430 Homo sapi
36	172.6	14.4	168880	9	AX017104	AX017104 Homo sapi
37	171	14.3	1360	6	AX451423	AX451423 Sequence
38	169.8	14.2	172978	2	BX530031	BX530031 Danio rer
39	169.8	14.2	233276	2	BX511175	BX511175 Danio rer
40	169.4	14.2	1134	6	AR202302	AR202302 Sequence
41	163	13.6	1191	6	AR217803	AR217803 Sequence
42	160	13.4	2003	6	AX513601	AX513601 Sequence
43	157.2	13.1	194363	5	AL954634	AL954634 Zebrafish
44	157.2	13.1	206683	2	BX323592	BX323592 Danio rer
45	156.8	13.1	1167	6	AR217813	AR217813 Sequence

ALIGNMENTS

RESULT 1
BC009075
LOCUS BC009075 2468 bp mRNA linear ROD 16-APR-2003
DEFINITION Mus musculus UDP-GlcNAc:betaGal
beta-1,3-N-acetylglucosaminyltransferase 1, mRNA (cDNA clone
MGC:6892 IMAGE:2654354), complete cds.
ACCESSION BC009075
VERSION BC009075.1 GI:14290591
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2468)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

Pred. No. is the number of results predicted by chance to have a

Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altschul, S.F., Zebberg, S., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickinson, K.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 2468)

Strausberg, R.

Direct Submission

Submitted (30-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NTI-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsoed, H., Kowis, C.R., Speed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAP Plate: 5 Row: e Column: 16

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 9938023, location/Qualifiers

1. 2468

/organism="Mus musculus"

/mol_type="mRNA"

/strain="NWRI"

/db_xref="taxon:10090"

/clone="MGC:6892 IMAGE:2654354"

/tissue_type="Mammary tumor. WAP-Tag model. 5 months old, gross tissue."

/clone_lib="NCI_CGAP_Mam4"

/lab_host="DH10B"

/note="vector: pCMV-SPORT6"

1. 2468

/gene="B3gnt1"

/note="synonyms: Beta3gnt, beta-3gnt, B3Galt6"

/db_xref="LocusID:53625"

/db_xref="MGI:1889505"

129.1322

/codon_start=1

/product="UDP-GlcNAc:betaGal

beta-1,3-N-acetylglucosaminyltransferase 1"

/protein_id="AAH09075.1"

/db_xref="GI:14290592"

FEATURES
source

gene

CDS

/db_xref="LocusID:53625"

/translation="MSVRRRVKLLGILMANVFYILVEVSKNSQDRKGGVILP
KEKFWKPPSPRAYWNRQELNRYNPILNRVKNGELATSPNTSLISYEPDSV
MTAVTDNNLPDRFKDFLLVLCRNTSLLDQPKKAKKPFLLATKSLIPHEARQA
IRESGRETWVGNOTVVRVELLCKTPEDNHPDLSMDLAFESDKHODILMNYRDFEF
NLSLKEVLPRLRWSTSCPDREYFVFKGDDDDVFVYTHLLNLSLSKSKAKDLFIGDVI
HNAGPHRDKLKIYIPEYFVPPYAGGGLFISGPLALRLYSKTSRVHLPIDDV
YTMCGLOKLGLVPEKHGKFTFDIEKNKNKNCISYIDLMLVHSRQEMIDWSQLQS
PNLKC"

BASE COUNT 685 a 509 c 564 g 710 t

ORIGIN

Query Match 100.0%; Score 1196; DB 10; Length 2468;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGGCGTCGAAGAGTCAAGTTGCTGGGCATCCTGATGATGGCAATGTCT 60
DB 127 AGATGAGTGTGGGGCGTCGAAGAGTCAAGTTGCTGGGCATCCTGATGATGGCAATGTCT 186

QY 61 TCATTTATTTGATTTGTGGAAGTCTCCAAAACAGTAGCCAAAGACAAAATGGAAGGGAG 120
DB 187 TCATTTATTTGATTTGTGGAAGTCTCCAAAACAGTAGCCAAAGACAAAATGGAAGGGAG 246

QY 121 GAGTAATAATCCCGAAAGAGAAGTTCTGGAAGCCACCAGCAGCTCCCGGGCAGTACTGGA 180
DB 247 GAGTAATAATCCCGAAAGAGAAGTTCTGGAAGCCACCAGCAGCTCCCGGGCAGTACTGGA 306

QY 181 ACAGGGACAGGAGAAGCTGAACAGTGTGACAAATCCCATCTTTGAACAGGGTGGCCAAATC 240
DB 307 ACAGGGACAGGAGAAGCTGAACAGTGTGACAAATCCCATCTTTGAACAGGGTGGCCAAATC 366

QY 241 AGACAGGGAGCTAGCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACACAGACT 300
DB 367 AGACAGGGAGCTAGCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACACAGACT 426

QY 301 CGAGGGTCATGACAGCTGTGACAGATTTTAAATCTGCCGGACAGATTTTAAAGACTTTC 360
DB 427 CGAGGGTCATGACAGCTGTGACAGATTTTAAATCTGCCGGACAGATTTTAAAGACTTTC 486

QY 361 TCTTGTATTTGAGATGCGGAATTTACTCGCTGCTTATAGATCAACGACAAATGTGCA 420
DB 487 TCTTGTATTTGAGATGCGGAATTTACTCGCTGCTTATAGATCAACGACAAATGTGCA 546

QY 421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTTCCACATTTTCCAGAGCAAG 480
DB 547 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTTCCACATTTTCCAGAGCAAG 606

QY 481 CAATTCGGAGTCTTGGGGCGGAAACCAACAGTAGGGAACACAGATAGTAGGAGTCT 540
DB 607 CAATTCGGAGTCTTGGGGCGGAAACCAACAGTAGGGAACACAGATAGTAGGAGTCT 666

QY 541 TCCTGTTGGGCAAGACACCCCGAGGACCAACACCTGACCTTTCCGACATCTTAAAGT 600
DB 667 TCCTGTTGGGCAAGACACCCCGAGGACCAACACCTGACCTTTCCGACATCTTAAAGT 726

QY 601 TTGAGAGTGACAAGCACACAGGACATCCTCATGTGGAACATATAGAGACACATTTCTCAACC 660
DB 727 TTGAGAGTGACAAGCACACAGGACATCCTCATGTGGAACATATAGAGACACATTTCTCAACC 786

QY 661 TGTCCCTGAAGGAAGTGTCTTCTTAGTGGGTGAGCAGCTTCTGTCAGAGCGCAGAGT 720
DB 787 TGTCCCTGAAGGAAGTGTCTTCTTAGTGGGTGAGCAGCTTCTGTCAGAGCGCAGAGT 846

QY 721 TTGTCTTCAAGGGCGATGATGACGCTGTTGTGGAACACCCATCATCTTAAATTAATCTGA 780
DB 847 TTGTCTTCAAGGGCGATGATGACGCTGTTGTGGAACACCCATCATCTTAAATTAATCTGA 906

QY 781 ATAGCTTATCCAAAGGACCAAGCCAAAGACTTGTTCATAGGTGAGCTGATCCCAATGCTG 840
DB 907 ATAGCTTATCCAAAGGACCAAGCCAAAGACTTGTTCATAGGTGAGCTGATCCCAATGCTG 966

QY 841 GGCCTCACCGGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACGGGGTCT 900

Db	967	GGCTCACCGGATAGAACTGAAGTACTACATCCAGAGTCTTCTACACGGGCTCT	1026
Qy	901	ACCACCGTATGCGGGGGTGGTGAATTCCTGTAATCCCGCCCGCTTGAGGCTGI	960
Db	1027	ACCACCGTATGCGGGGGTGGTGAATTCCTGTAATCCCGCCCGCTTGAGGCTGI	1086
Qy	961	ACAGTGGGACTAGCGGGTCCATCTCTACCTATTGATGATGATGATGATGATGATG	1020
Db	1087	ACAGTGGGACTAGCGGGTCCATCTCTACCTATTGATGATGATGATGATGATGATG	1146
Qy	1021	TTCAAGAACTGGGCTTGTTCAGAGAGACACAAAGCTTCAGGACATTTGATATTGAAG	1080
Db	1147	TTCAAGAACTGGGCTTGTTCAGAGAGACACAAAGCTTCAGGACATTTGATATTGAAG	1266
Qy	1081	AGAAATAAGAAAAATATTGTTCCCTATATAGACCTAATGTTAGTACATAGCAGAAAC	1140
Db	1207	AGAAATAAGAAAAATATTGTTCCCTATATAGACCTAATGTTAGTACATAGCAGAAAC	1266
Qy	1141	CTCAAGAGATGATTGATATCTGGTCTCAGTTGCAAGCTCCTAATTTAAATGCTGA	1196
Db	1267	CTCAAGAGATGATTGATATCTGGTCTCAGTTGCAAGCTCCTAATTTAAATGCTGA	1322
RESULT 2			
LOCUS	AY043479	2672 bp mRNA linear	ROD 02-SEP-2001
DEFINITION	Mus musculus beta-1,3-N-acetylglucosaminyltransferase mRNA, complete cds.		
ACCESSION	AY043479		
VERSION	AY043479.1	GI:15421159	
KEYWORDS			
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2672)		
AUTHORS	Wang, S.E., Cohen, B.L., Sarkar, M., Ying, Y., Cohen, S., Singh, N., Wang, W., Flock, G., Goh, T. and Schachter, H.		
TITLE	Molecular cloning and expression analysis of a mouse UDP-GlcNAc:Gal(beta1-4)Glc(NAc)-R		
JOURNAL	beta1,3-N-acetylglucosaminyltransferase homologous to Drosophila melanogaster Brainiac and the beta1,3-galactosyltransferase family Glycoconj. J. 17, 865-872 (2000)		
REFERENCE	2 (bases 1 to 2672)		
AUTHORS	Egan, S.E., Cohen, B.L., Sarkar, M., Ying, Y., Cohen, S., Singh, N., Wang, W., Flock, G., Goh, T. and Schachter, H.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-JUL-2001) Programs in Cancer Research and Developmental Biology, The Hospital for Sick Children, 555 University Avenue, Toronto, ON M5G 1X8, Canada		
FEATURES	Location/Qualifiers		
source	1..2672		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="C.B17"		
	/db_xref="taxon:10090"		
	210..1403		
	/note="b3gnf"		
	/codon_start=1		
	/product="beta-1,3-N-acetylglucosaminyltransferase"		
	/protein_id="AAK95359.1"		
	/db_xref="gi:15421160"		
CDS			
	/translation="MSVGRRRVLLGLMNVNFIYLIVESKNSQDKNGKGVLIIP KEATVDNPPDRKDELLYLRNYSLLIDQPKKAKKPFLLALSLIPHFAROA IRESWRETNQVTVRVFLGKTPPEDNHPDLSMKLFEESDKHODILMWNRYDTFF NLSLKEVLFWRVSTGPDFAEFVKFGDDDFVFNTHILNLSLSKSKAKDLFGVY HNAGPRDKLKYVIPVFTGYVPPVAGGGGFLYSGPLALRYLSATSRVILYPIDV YTGKCLQKLGILVPEKHKGFRTFDEENKKNKCSYIDLMLVHSRKKPOEMIDVNSLQS PNLKC"		
BASE COUNT	731 a 562 c 635 g 744 t		
ORIGIN			

Query Match	100.0%	Score 1196;	DB 10;	Length 2672;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1196;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	AGATGAGTGTGGGGCTCGAAGAGTCAAGTCTCTGGGATCTCTGATGATGGCAAAATGTCT	60	
Db	208	AGATGAGTGTGGGGCTCGAAGAGTCAAGTCTCTGGGATCTCTGATGATGGCAAAATGTCT	267	
Qy	61	TCATTATTATTGATTTGGGAAGTCTCCAAAACAGTAGCCAAAGCAAAAATGAAAGGGAG	120	
Db	268	TCATTATTATTGATTTGGGAAGTCTCCAAAACAGTAGCCAAAGCAAAAATGAAAGGGAG	327	
Qy	121	GAGTATAATCCGAAAGAGAGATTCTGGAAGCCACCAGCACTCCCGGGCATACTGGA	180	
Db	328	GAGTATAATCCGAAAGAGAGATTCTGGAAGCCACCAGCACTCCCGGGCATACTGGA	387	
Qy	181	ACAGGGAACAGGAGAGCTGAACAGGTGTAACATCCCATCTTTGAACAGGTGGCCAATC	240	
Db	388	ACAGGGAACAGGAGAGCTGAACAGGTGTAACATCCCATCTTTGAACAGGTGGCCAATC	447	
Qy	241	AGACAGGGAGCTAGCCACATCTCCAAACAGTACCTGAGCTATTGTGAACAGACACT	300	
Db	448	AGACAGGGAGCTAGCCACATCTCCAAACAGTACCTGAGCTATTGTGAACAGACT	507	
Qy	301	CGACGGTCATGACAGCTGTGACAGATTTTAAATAATCTGCCGACAGATTAAAGACTTC	360	
Db	508	CGACGGTCATGACAGCTGTGACAGATTTTAAATAATCTGCCGACAGATTAAAGACTTC	567	
Qy	361	TCTTGATTGTGAGTCCCGGAATTAATCTGCTCTTATAGATCAACGAGAAATGTGAA	420	
Db	568	TCTTGATTGTGAGTCCCGGAATTAATCTGCTCTTATAGATCAACGAGAAATGTGAA	627	
Qy	421	AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCATTCACATTTTGGCCAGAGCAAG	480	
Db	628	AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCATTCACATTTTGGCCAGAGCAAG	687	
Qy	481	CAATTCCGGAGTCTTGGGGCCGAGAAACAGCTAGGGAACAGACAGTAGTGAGGCTT	540	
Db	688	CAATTCCGGAGTCTTGGGGCCGAGAAACAGCTAGGGAACAGACAGTAGTGAGGCTT	747	
Qy	541	TCCTGTGGGCAAGACACCCAGAGCAACACCCCTGACCTTCCTGGACATGCTTAAGT	600	
Db	748	TCCTGTGGGCAAGACACCCAGAGCAACACCCCTGACCTTCCTGGACATGCTTAAGT	807	
Qy	601	TTGAGAGTGACAAAGCAACAGGACATCTCATGTGGAACATATAGACACATTTCTCAACC	660	
Db	808	TTGAGAGTGACAAAGCAACAGGACATCTCATGTGGAACATATAGACACATTTCTCAACC	867	
Qy	661	TGTCCTCTGAAGGAAGTCTTCTTAGTGGTGAGCACTTCTCTCCAGACGAGAGT	720	
Db	868	TGTCCTCTGAAGGAAGTCTTCTTAGTGGTGAGCACTTCTCTCCAGACGAGAGT	927	
Qy	721	TTGTCCTTCAAGGGCGATGATGACGTGTTGTGAACACCCATCACATCTTAATTTACTTGA	780	
Db	928	TTGTCCTTCAAGGGCGATGATGACGTGTTGTGAACACCCATCACATCTTAATTTACTTGA	987	
Qy	781	ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGGTACGTGATCCACATGCTG	840	
Db	988	ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGGTACGTGATCCACATGCTG	1047	
Qy	841	GGCTCACCGGATAGAACTGAAGTACTACATCCAGAACTCTTCTACACGGGCTCT	900	
Db	1048	GGCTCACCGGATAGAACTGAAGTACTACATCCAGAACTCTTCTACACGGGCTCT	1107	
Qy	901	ACCACCGTATGCGGGGGTGGTGAATTCCTGTACTCCGGCCCGCTTGAGGCTGT	960	
Db	1108	ACCACCGTATGCGGGGGTGGTGAATTCCTGTACTCCGGCCCGCTTGAGGCTGT	1167	
Qy	961	ACAGTGGGACTAGCGGGTCCATCTCTACCTATTGATGATGATGATGATGATGATG	1020	
Db	1168	ACAGTGGGACTAGCGGGTCCATCTCTACCTATTGATGATGATGATGATGATGATG	1227	

QY 1021 TTCAGAACTGGGCTTCTTCAGAGAGACAGAAAGCTTCAGACATTTGATTAAG 1080
|||||
Db 1228 TTCAGAACTGGGCTTCTTCAGAGAGACAGAAAGCTTCAGACATTTGATTAAG 1287
|||||
QY 1081 AGAAAAATAGAAAAATATTTGCTTATATAGACCTTAATGTTAGTACATAGCAGAAAC 1140
|||||
Db 1288 AGAAAAATAGAAAAATATTTGCTTATATAGACCTTAATGTTAGTACATAGCAGAAAC 1347
|||||
QY 1141 CTCAGAGATGATGATATCTGGTCTCAGCTGCAAGTCTCTAATTTAAATGCTGA 1196
|||||
Db 1348 CTCAGAGATGATGATATCTGGTCTCAGCTGCAAGTCTCTAATTTAAATGCTGA 1403
|||||

RESULT 3

AL772364/c 210863 bp DNA linear ROD 18-FEB-2003
LOCUS Mouse DNA sequence from clone RP23-242C19 on chromosome 11,
DEFINITION complete sequence.

AL772364

AL772364

HTG.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 210863)

Andrew.R.

Direct Submission

Submitted (18-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

On Feb 18, 2003 this sequence version replaced g1:28300672.

Sequence from the Mouse Genome Sequencing Consortium whole genome

shotgun may have been used to confirm this sequence. Sequence data

from the whole genome shotgun alone has only been used where it has

a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest, except on the rare

occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession:

numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; WP., WormPEP; Information

on the WormPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep

from the RPI-23 Mouse BAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6.

Location/Qualifiers

1..210863

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="11"

/clone="RP23-242C19"

/clone_lib="RPI-23"

FEATURES

source

BASE COUNT	62433 a	47427 c	45742 g	55261 t	ORIGIN
Query Match	100.0%	Score 1196;	DB 10;	Length 210863;	
Best Local Similarity	100.0%	Pred. No. 0;			
Matches 1196;	Conservative	0;	Mismatches	0;	Indels
0;	Gaps	0;			
QY 1	AGATGAGTGTGGGGTTCGAAGAGTCAAGTTGCTGGGCATCCTGATGATGGCAATGCT	60			
Db 71238	AGATGAGTGTGGGGTTCGAAGAGTCAAGTTGCTGGGCATCCTGATGATGGCAATGCT	71179			
QY 61	TCATTATTTCATTGTCGAAGTCTCCAAACACAGTACCCAGACAAATATGGAAGGAG	120			
Db 7-178	TCATTATTTCATTGTCGAAGTCTCCAAACACAGTACCCAGACAAATATGGAAGGAG	71119			
QY 121	GAGTAATAATCCGAAAGAGAAAGTTCTGGAAGCCACCCAGACATCCCGGGCATACTGGA	180			
Db 71118	GAGTAATAATCCGAAAGAGAAAGTTCTGGAAGCCACCCAGACATCCCGGGCATACTGGA	71059			
QY 181	ACAGGGAACAGAGAGAGTGAACAGGTGGTACATCCATCTTGAACAGGGTGCCCAATC	240			
Db 71058	ACAGGGAACAGAGAGAGTGAACAGGTGGTACATCCATCTTGAACAGGGTGCCCAATC	70999			
QY 241	AGACAGGGGAGCTAGCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACT	300			
Db 70998	AGACAGGGGAGCTAGCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACT	70939			
QY 301	CGACGGTCTATGACAGCTGTGACAGATTTTAATAATCTGCCGGACAGATTTAAAGACTTC	360			
Db 70938	CGACGGTCTATGACAGCTGTGACAGATTTTAATAATCTGCCGGACAGATTTAAAGACTTC	70879			
QY 361	TCTTGTATTGAGATGCGGGAATTAATGCTGCTTATAGATCAACCGAAGAAATGTGCA	420			
Db 70878	TCTTGTATTGAGATGCGGGAATTAATGCTGCTTATAGATCAACCGAAGAAATGTGCA	70819			
QY 421	AGAAGCCCTTCTTACTATTGGGATAAAGTCCCTCATTCACATTTTGCAGAGGGAAG	480			
Db 70818	AGAAGCCCTTCTTACTATTGGGATAAAGTCCCTCATTCACATTTTGCAGAGGGAAG	70759			
QY 481	CAATTTCGGGAGTCTTGGGGCGGAGAACCAACAGTAGGGAACACACAGTAGTGGGTCT	540			
Db 70758	CAATTTCGGGAGTCTTGGGGCGGAGAACCAACAGTAGGGAACACACAGTAGTGGGTCT	70699			
QY 541	TCCTGTTGGGGAAGACACCCCGGAGGACAAACACCTGACCTTCGGACATGCTTAAGT	600			
Db 70698	TCCTGTTGGGGAAGACACCCCGGAGGACAAACACCTGACCTTCGGACATGCTTAAGT	70639			
QY 601	TTGAGAGTGACAAGCACACAGACATCCTCATGTGGAACATATAGACACATTTCTCAAC	660			
Db 70638	TTGAGAGTGACAAGCACACAGACATCCTCATGTGGAACATATAGACACATTTCTCAAC	70579			
QY 661	TGTCCTTGAAGGAAGTGTGTTTCTTAGTGGGTGAGCAGCTTCCTGCCAGACGACAGT	720			
Db 70578	TGTCCTTGAAGGAAGTGTGTTTCTTAGTGGGTGAGCAGCTTCCTGCCAGACGACAGT	70519			
QY 721	TTGCTTCAAGGGGATGACAGCTGTTTGTGAACACCCATCATCTTAAATTAATGTA	780			
Db 70518	TTGCTTCAAGGGGATGACAGCTGTTTGTGAACACCCATCATCTTAAATTAATGTA	70459			
QY 781	ATAGCTTATCCAAGAGCAAGCCCAAGACTTGTTCATAGGTGAGTGCATCCACAATGCT	840			
Db 70458	ATAGCTTATCCAAGAGCAAGCCCAAGACTTGTTCATAGGTGAGTGCATCCACAATGCT	70399			
QY 841	GSCCTTACCGGATTAAGAACTGAAGTACTACATCCAGAGTCTTCTACACCGGCTCT	900			
Db 70398	GSCCTTACCGGATTAAGAACTGAAGTACTACATCCAGAGTCTTCTACACCGGCTCT	70339			
QY 901	ACCCACCGTATCCCGGGGTGGTGGATTCCTGTACTCCGGCCCCCTTGCCTTGAGGCTGT	960			
Db 70338	ACCCACCGTATCCCGGGGTGGTGGATTCCTGTACTCCGGCCCCCTTGCCTTGAGGCTGT	70279			
QY 961	ACAGTGCACACTAGCCGGGTCCATCTTACCCCTATTGATGATGTTTACGGGAATGTGCC	1020			

Db	70278	ACAGIGGAGCTAGCGGGGCCATCTCTACCCATTATGATGATGTTATACGGGATGTGCC	70219
Qy	1021	TTGAGAACTGGGCCTTGTTCAGAGAGACAAAGGCTTCAGGACATTGATATTGAAG	1080
Db	70218	TTGAGAACTGGGCCTTGTTCAGAGAGACAAAGGCTTCAGGACATTGATATTGAAG	70159
Qy	1081	AGAAATATAGAAATATTTGTTCTCTATATAGACCTAATGTTAGTACATAGCAGAAAC	1140
Db	70158	AGAAATATAGAAATATTTGTTCTCTATATAGACCTAATGTTAGTACATAGCAGAAAC	70099
Qy	1141	CTCAAGAGATGATGATATCTGCTCAGTTCGCAAGTCTCAATTTAAATGCTGA	1196
Db	70098	CTCAAGAGATGATGATATCTGCTCAGTTCGCAAGTCTCAATTTAAATGCTGA	70043
RESULT 4			
AF092050			
LOCUS		1260 bp mRNA linear ROD 09-AUG-2000	
DEFINITION		Mus musculus beta-1,3-N-acetylglucosaminyltransferase mRNA.	
ACCESSION		AF092050	
VERSION		AF092050.2 GI:9755418	
KEYWORDS			
SOURCE		Mus musculus (house mouse)	
ORGANISM		Mus musculus	
REFERENCE			
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE		1. (bases 1 to 1260)	
JOURNAL		Zhou,D., Dinter,A., Gutierrez Gallego,R., Kamberling,J.P., Vilegandhant,J.F., Berger,E.G. and Hennet,T.	
MEDLINE		A beta-1,3-N-acetylglucosaminyltransferase with	
PUBMED		poly-N-acetylglucosamine synthase activity is structurally related	
REFERENCE		to beta-1,3-galactosyltransferases	
AUTHORS		Proc. Natl. Acad. Sci. U.S.A. 96 (2), 406-411 (1999)	
TITLE		99110903	
JOURNAL			
MEDLINE		2. (bases 1 to 1260)	
PUBMED		Zhou,D., Berger,E.G. and Hennet,T.	
REFERENCE		Direct Submission	
AUTHORS		Submitted (15-SEP-1998) Physiology, University of Zurich,	
TITLE		Winterthurerstrasse 190, Zurich 8057, Switzerland	
JOURNAL		3. (bases 1 to 1260)	
MEDLINE		Zhou,D., Berger,E.G. and Hennet,T.	
PUBMED		Submitted (09-AUG-2000) Physiology, University of Zurich,	
REFERENCE		Winterthurerstrasse 190, Zurich 8057, Switzerland	
AUTHORS		Sequence update by submitter	
TITLE		On Aug 9, 2000 this sequence version replaced gi:4191391.	
JOURNAL		Location/Qualifiers	
REMARK		1. .1260	
COMMENT		/organism="Mus musculus"	
FEATURES		/mol_type="mRNA"	
source		/db_xref="taxon:10090"	
		/tissue_type="brain"	
		/dev_stage="newborn"	
		30. .1223	
		/EC_number="2.4.1.149"	
		/note="glycosyltransferase; b3GnT"	
		/codon_start=1	
		/product="beta-1,3-N-acetylglucosaminyltransferase"	
		/protein_id="AAB09763.2"	
		/db_xref="GI:9755419"	
		/translation="MSVRRRVKLLGLIMANVFYLYIVEVSKNSQDKNGGVIIIP	
		KERFAPSPRAYNREDEKLNRYNPILNRVANOTGELATSPNTHSLSYCEPDSTV	
		IRSVGRETVNGQTVYVFLGKTPEDNPDLSMLKFDSDKHQDILMWNRYRTFF	
		NLSLKEVFLRWSTSCPAEFVKGDVDVFNTHILNLSLSKSKAKDLF IGDI	
		HNAGPHSDKKLYI IPEVFTGVYPPYAGGGFLYSPPLALPLYSATSRVHLYPIDDV	
		YTCMLQKLGVLPEKHGFRFIDIEKNKNKNCYSIDLMLVHSRKPQEMIDVWSLQCS	
		PNLKC.	
BASE COUNT		370 a 282 c 307 g 301 t	
ORIGIN			

Query Match	99.7%	Score 1192.8	DB 10	Length 1260
Best Local Similarity	99.8%	Pred. No. 0		
Matches 1194	Conservative	0	Mismatches	2
			Indels	0
			Gaps	0
QY	1	AGATCAGTGTGGGGGTGCGAAGAGTCAAGTTGCTGGGCATCCTGATGATGCAAAATGCT	60	
DB	28	AGATCAGTGTGGGGGTGCGAAGAGTCAAGTTGCTGGGCATCCTGATGATGCAAAATGCT	87	
QY	61	TCATTATTATTGATTGGAAGTCTCCAAAACAGTAGCAAGACAAAATGGAAGGAG	120	
DB	88	TCATTATTATTGATTGGAAGTCTCCAAAACAGTAGCAAGACAAAATGGAAGGAG	147	
QY	121	GAGTAAATATCCGAAAGAGAGTTCTGGAAGCCACCAGCACTCCCGGGCATACTGGA	180	
DB	148	GAGTAAATATCCGAAAGAGAGTTCTGGAAGCCACCAGCACTCCCGGGCATACTGGA	207	
QY	181	ACAGGGAACAGGAGAGCTGAACAGTGTCAATCCCATCTTGAACAGGTGGCCAAATC	240	
DB	208	ACAGGGAACAGGAGAGCTGAACAGTGTCAATCCCATCTTGAACAGGTGGCCAAATC	267	
QY	241	AGACAGGGAGCTAGCCACATCTCCAAAACACAAGTCACTGAGCTATTGTGAACAGACT	300	
DB	268	AGACAGGGAGCTAGCCACATCTCCAAAACACAAGTCACTGAGCTATTGTGAACAGACT	327	
QY	301	CGACCGTCATGACAGCTGTGACAGATTTTAATAATCTCCGGGACAGATTAAAGACITTC	360	
DB	328	CGACCGTCATGACAGCTGTGACAGATTTTAATAATCTCCGGGACAGATTAAAGACITTC	387	
QY	361	TCCTGATTGAGATGCCGGAAATTAATCGCTTATAGATCAACCGAAGAAATGTGAA	420	
DB	388	TCCTGATTGAGATGCCGGAAATTAATCGCTTATAGATCAACCGAAGAAATGTGAA	447	
QY	421	AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCAITCCACATTTTCCCAAGAGGCAAG	480	
DB	448	AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCAITCCACATTTTCCCAAGAGGCAAG	507	
QY	481	CAATTCCGGAGTCTTGGGGCGGAGAACCAACAGTAGGGAACACAGACAGTACTGAGGCT	540	
DB	508	CAATTCCGGAGTCTTGGGGCGGAGAACCAACAGTAGGGAACACAGACAGTACTGAGGCT	567	
QY	541	TCCTGTTGGGCAAGACACCCCGAGAGCAACCCCTGACCTTTCGCACATGCTTAAGT	600	
DB	568	TCCTGTTGGGCAAGACACCCCGAGAGCAACCCCTGACCTTTCGCACATGCTTAAGT	627	
QY	601	TTGAGAGTGACAAGCACCAGGACATCTCATGTGGAACATATAGACACACATCTTCAACC	660	
DB	628	TTGAGAGTGACAAGCACCAGGACATCTCTCATGTGGAACATATAGACACACATCTTCAACC	687	
QY	661	TGTCCTTGAAGGAAGTGTGTTCTTAGTGGTGAGCACTTCTGTCACACACAGAGT	720	
DB	688	TGTCCTTGAAGGAAGTGTGTTCTTAGTGGTGAGCACTTCTGTCACACACAGAGT	747	
QY	721	TTGCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCAGTCTTAATTAATTGGA	780	
DB	748	TTGCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCAGTCTTAATTAATTGGA	807	
QY	781	ATAGCTTATCCAAAGAGAAAGCCAAAGACTTGTTCATAGGTGAGTGTATCCACAAATGCTG	840	
DB	808	ATAGCTTATCCAAAGAGAAAGCCAAAGACTTGTTCATAGGTGAGTGTATCCACAAATGCTG	867	
QY	841	GGCCTCACCGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACCGGCTCT	900	
DB	868	GGCCTCACCGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACCGGCTCT	927	
QY	901	ACCCACCGTATGCCGGGGGTGTGGATTCTCTGACTCCGGCCCCCTTGCCTTTAGGGTGT	960	
DB	928	ACCCACCGTATGCCGGGGGTGTGGATTCTCTGACTCCGGCCCCCTTGCCTTTAGGGTGT	987	
QY	961	ACAGTGCAGCTAGCCGGGTCCATCTCTCCCTATTGATGATGTTTATACGGGAATGTGCC	1020	
DB	988	ACAGTGCAGCTAGCCGGGTCCATCTCTCCCTATTGATGATGTTTATACGGGAATGTGCC	1047	

Q7 1021 TTCAGAACTGGCCCTTTCACAGAGACAGACAAAGCTTCAGGACATTTGATATTGAAG 1080
 |||||||
 Db 1048 TTCAGAACTGGCCCTTTCACAGAGACAGACAAAGCTTCAGGACATTTGATATTGAAG 1107
 |||||||
 Q7 1081 AGAAATAAAGAAATATTTGCTCTATACACCTAACTGTTAGTACATACAGAAAC 1140
 |||||||
 Db 1108 AGAAATAAAGAAATATTTGCTCTATACACCTAACTGTTAGTACATACAGAAAC 1167
 |||||||
 Q7 1141 CTCAGAGATGATGATATCTGCTCAGTTCGAAAGTCTCTAATTTAAATCTGA 1196
 |||||||
 Db 1168 CTCAGAGATGATGATATCTGCTCAGTTCGAAAGTCTCTAATTTAAATCTGGA 1223
 |||||||

RESULT 5
 AC119549/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-444D1, WORKING DRAFT SEQUENCE, 2
 unordered pieces.
 AC119549
 AC119549.5 GI:25008693
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 178009)
 Muzny,D.,Marlie,, Metzker,M.,Lee,, Abramson,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
 Georgergis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hayes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
 Hollins,B., Howells,S., Hui,Y., Hume,J., Iglebird,D., Jackson,A.,
 Jackson,L., Jacob,S., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Loresuhewa,L., Loulsegod,H., Lozado,R., Lu,X., Ma,J.,
 Maheshwari,M., Mahindaratne,M., Mahmoud,M., Mailloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,M., Martinez,E.,
 Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
 Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
 Pasternak,S., Paul,H., Perez,A., Perez,I., Prannkoch,C.,
 Piopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
 Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
 Shetty,J., Shvartsdeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
 Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
 Williams,G., Willson,R., Wleczka,R., Wooden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinstock,G. and Gibbs,R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 178009)
 Worley,K.C.
 Direct Submission
 Submitted (28-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 178009)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 15, 2002 this sequence version replaced gi:23194601.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GMD0
 Center clone name: CH230-444D1
 ----- Summary Statistics
 Assembly program: Phrap: version 0.990329
 Consensus quality: 156182 bases at least Q40
 Consensus quality: 157889 bases at least Q30
 Consensus quality: 159078 bases at least Q20
 Estimated insert size: 161372; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 175629: contig of 175629 bp in length
 * 175630 175729: gap of unknown length
 * 175730 178009: contig of 2280 bp in length.
 FEATURES
 source
 location/Qualifiers
 1..178009
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-444D1"
 785..1598
 /note="clone_boundary
 clone_end:T7
 site:
 end_sequence:BZ198810"
 complement(166970..167869)
 /note="clone_boundary
 clone_end:Sp6
 misc_feature
 785..1598
 /note="clone_boundary
 clone_end:T7
 site:
 end_sequence:BZ198810"
 complement(166970..167869)
 /note="clone_boundary
 clone_end:Sp6
 misc_feature
 785..1598
 /note="clone_boundary
 clone_end:T7
 site:
 end_sequence:BZ198810"
 complement(166970..167869)
 /note="clone_boundary
 clone_end:Sp6

Db	143835	GGCCCCATCGGGATAAGAAACTGAAGTACTACATCCCGAAGTCTTCTTACACTGGGGTCT	143777
QY	901	ACCCACCGGTATCGCGGGGTGGGGATTCCTGTACTCCGGCCCGCTTGCCTTCAGGCTGT	950
Db	143775	ACCCACCGTATCGCGGGGTGGGGATTCCTGTACTCCGGCCCGCTTGCCTTCAGGCTTT	143716
QY	961	ACAGTGGGACTAGCGGGGTCCATCTTACCCCTATTGATGATGTTTATACGGGAATGGC	1020
Db	143715	ACAAATGTAACTACCGCGGTCCACCTCTACCCCATCGATGACGTTTATCTAGGAATGGC	143656
QY	1021	TTTCAGAACTGGCGCTTCTTCAGAGAGACACAAGGCTTCAGGACATTTGATATTGAAS	1080
Db	143655	TTTCAGAACTGGCGCTTCTTCAGAGAGACACAAGGCTTCAGGACATTTGATATTGAAS	143596
QY	1081	AGAAAAATAAGAAAAATATTTGTTCTATATAGACCTTAATGTTAGTACATAGCAGAAAC	1140
Db	143595	AGAAAAATAAGAAAAATATTTGTTCTATATAGATCTTAATGTTAGTACACAGCAGAAAC	143536
QY	1141	CTCAGAGATGATTGATATCTGGTCTCAGTTGCAAGTCTCTAATTTAAATGCT	1194
Db	143535	CTCAGAGATGATTGATATCTGGTCTCAGTTGCAAGTCTCTAATTTAAATGCT	143482
RESULT	6		
AC109547			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			

Db	300726	TTGCTTTCAGGGTGATGATGAGCTGTCTGTAACCCATACATCTCTGAATACITGA	300785
Qy	781	ATAGCTTATCCAGAGCAAGCAAGACTTGTTCATAGTGACGTGATCCACAAATGCTG	840
Db	300786	ATAGCTTATCCAGAGCAAGCAAGACTTGTTCATAGTGACGTGATCCACAAATGCTG	300845
Qy	841	GGCTCACCGGATAGCAAACTGAAGTACTACATCCAGAGTCTTCTACACCGGGTCT	900
Db	300846	GGCCCCATCGGGATAGCAAACTGAAGTACTACATCCAGAGTCTTCTACACTGGGTCT	300905
Qy	901	ACCACCGTATGCCGGGGTGGTGATTCCTGTACTCCGGCCCCCTTGCCTTACAGGGTGT	960
Db	300906	ACCACCGTATGCCAGGGGGTGGGTTCCTGTACTTGGCCCCGCTTACAGGGCTTT	300965
Qy	961	ACAGTGGCATAGCCGGGTGCATCTCTACCCATTTGATGATGTTTATACGGGAATGTC	1020
Db	300966	ACATGTAAGTACCGGGTGCACCTCTACCCATTCGATTAACGCTCAATCTGGAATGTC	301025
Qy	1021	TTCAGAACTGGGCTTGTTCAGAGCAAGCAAGGCTTCAGACATTTGATATGAAG	1080
Db	301026	TTCAGAACTGGGCTTGTTCAGAGCAAGCAAGGCTTCAGACATTTGATATGAAG	301085
Qy	1081	AGAAAAATAGAAAAATATTTGTTCTATATAGACCTAAAGTTAGTACATAGCAAGAAC	1140
Db	301086	AGAAAAATAGAAAAATATTTGTTCTATATAGATCTAATGTTAGTACATAGCAAGAAC	301145
Qy	1141	CTCAGAGATGATGATATCTGCTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGT	1194
Db	301146	CTCAGAGATGATGATATCTGCTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGT	301199
RESULT 7			
AF092051	1831 bp mRNA lineas: PRI 09-AUG-2000		
LOCUS	Homo sapiens beta-1,3-N-acetylglucosaminyltransferase mRNA,		
DEFINITION	complete cds.		
ACCESSION	AF092051		
VERSION	AF092051.2 GI:9755416		
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1831)		
	Zhou,D., Dinter,A., Gutierrez Gallego,R., Kamerling,J.P.,		
JOURNAL	Viegandhart,J.F., Berger,E.G. and Hennet,T.		
	A beta-1,3-N-acetylglucosaminyltransferase with		
MEDLINE	poly-N-acetylglucosamine synthase activity is structurally related		
	to beta-1,3-galactosyltransferases		
PUBMED	Proc. Natl. Acad. Sci. U.S.A. 96 (2), 406-411 (1999)		
REFERENCE	99110903		
AUTHORS	2 (bases 1 to 1831)		
TITLE	Zhou,D., Berger,E.G. and Hennet,T.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (15-SEP-1998) Physiology, University of Zurich,		
	Winterthurerstrasse 190, Zurich 8057, Switzerland		
AUTHORS	3 (bases 1 to 1831)		
TITLE	Zhou,D., Berger,E.G. and Hennet,T.		
JOURNAL	Direct Submission		
REMARK	Submitted (09-AUG-2000) Physiology, University of Zurich,		
	Winterthurerstrasse 190, Zurich 8057, Switzerland		
COMMENT	Sequence update by submitter		
FEATURES	On Aug 9, 2000 this sequence version replaced gi:4191393.		
source	Location/Qualifiers		
	1..1831		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/tissue_type="brain"		
	/dev_stage="fetus"		
	236..1429		
CDS			

/EC_number="2.4.1.149"			
/note="glycosyltransferase; b3Gnt"			
/codon_start=1			
/product="beta-1,3-N-acetylglucosaminyltransferase"			
/protein_id="AAD09764.2"			
/db_xref="GI:9755417"			
/translation="MSVRRRIKLLILMMANVFIYFIMEVSKSSQKNGKEVIIP			
KEFWKISTIPPEAYNROEKLRNYNPILSMLTNOTGEARLSNLSHNTCEPLRV			
TSVVTGFNNLPORFDLLYLCRNYSLIDQDKCAKKPFLLLAKSLTFHARRQA			
IESGQESNAGNQTVPVFLGQTPPDNDHPDLGMLKFESEKHODILMNNYRDTFF			
NLSLKEVLPFLRWSTSCDPTFEVKGDDDFVETHIILNLSLKTAKDLFDGDI			
HNAGPHRDKLKYIPEVYVSGLYPEVAGGGFLYSGHLALRLYHTDOVHLYPDDV			
YTGMLCKLGLVPEKHKGFTEDIEEKNKNNICSYVDLMLVHSRPEQEMIDINSLQ			
AHLKC"			
BASE COUNT	497 a	423 c	443 g 468 t
ORIGIN			
Query Match	75.8%; Score 906.4; DB 9; Length 1831;		
Best Local Similarity	84.9%; Pred. No. 1.5e-244;		
Matches :015; Conservative	0; Mismatches 181; Indels 0; Gaps 0;		
Qy	1	AGATGAGTGTGGGGGTGCAAGAGTCAAGTGTCTGGGCATCCTGATGATGGCAATGTCT	60
Db	234	AAATGAGTGTGGACGTCCAGAAATAAGTGTGTGGGTATCTGATGATGGCAATGTCT	293
Qy	61	TCATTATTGATGTGGAGTCTCCAAAACAGTAGCCAAACAAAATGGAAGGGAG	120
Db	294	TCATTATTATTATGGAAGTCTCCAAAAGCAGTAGCCAAAATAATGGAAGGGG	353
Qy	121	GAGTAATAATCCGGAAGAGAGTCTGGAAGCCACCCAGCAGTCCCGGGGCATCTGGA	180
Db	354	AGTAATAATCCCAAGAGAGTCTGGAAGATATCTACCCCTCCCGAGGCACTGGA	413
Qy	181	ACAGGGAACAGAGAGTGAACAGGTGGTACAAATCCCATCTTGAACAGGGTGGCAATC	240
Db	414	ACCGAGCAAGAGAGAGTGAACCGGCAGTACAAACCCCATCTGAGCAGTCTGACCAAC	473
Qy	241	AGACAGGGGAGTGCACATCTCCAAACAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA	300
Db	474	AGACAGGGGAGGGGGGAGGCTCTCCAAATATTAAGCCATCTGAACCTACTCCGACCT	533
Qy	301	CGACGGTCAATGACAGCTGTGACAGATTTTAAATAATCTCGCGGACAGATTTAAAGAC	360
Db	534	TGAGGGTCAAGTGGTGTAGGGTTTAAACAACTTCCCGGACAGATTTAAAGACTTTC	593
Qy	361	TCCTGTATTGAGATCCCGGAATTTACTCGCTGCTTATAGATCAACCGAAGAAATGTGAA	420
Db	594	TGCTGTATTGAGATCCCGCAATTTATTCACCTGCTTATAGATCAGCGGAGATAAGTGC	653
Qy	421	AGNAGCCCTTCTTACTATTGGGGATAAAGTCCCTCATTCACATTTTGGCAGAGCAAG	480
Db	654	AGAAACCTTTCTTGTGTGGGATTAAGTCCCTCACTCCACATTTTGGCAGAGCAAG	713
Qy	481	CAATTCGGGAGTCTTGGGGCCGAGAAACCAACGTTAGGGAACCAAGACAGTAGTGGGT	540
Db	714	CAATTCGGGAATCTTGGGGCCGAGAAACCAACGTTAGGGAACCAAGACAGTAGTGG	773
Qy	541	TCCTGTGGGCAAGACACCCCGAGGCAACACCCCTGACCTTTTGGACATCTTAAAGT	600
Db	774	TCCTGTGGGCAAGACACCCCGAGGCAACACCCCTGACATATGCTGAAT	833
Qy	601	TTGAGAGTCAAGACACCCCGAGGCAACCTCATGTGAACATATAGAGACATCTTCAAC	660
Db	834	TTGAGAGTCAAGACACCCCGAGGCAACCTCATGTGAACATATAGAGACATCTTCAAC	893
Qy	661	TGTCCTGAAGCAAGTGTGTTTCTTAGTGGGTGAGCACTTCCCTGTCAGAGCAGGAGT	720
Db	894	TGTCCTGAAGCAAGTGTGTTTCTTAGTGGGTGAGCACTTCCCTGTCAGAGCAGGAGT	953
Qy	721	TTGCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCCATCACATCTTAAATTTACTTGA	780
Db	954	TTGCTTCAAGGGCGATGACGTGTTTGTGAACACCCCATCACATCTTAAATTTACTTGA	1013

QY	781	ATAGCTTATCCAGAGAGCAAGCCAAAGACTTGITCATAGGTGACGTGATCCACAAAGCTG	843
Db	1014	ATAGTTTATCCAGAGCAAGCCAAAGACTTCTTCATAGGTGATGATGATCCACAAATGCTG	1073
QY	841	GGCTCACCAGGTAAGCAAACTGAAGTACTACATCCACAGAGCTTCTACACGGCGTCT	900
Db	1074	GACCTATCGGTAAGCAAACTGAAGTACTACATCCACAGAGCTTCTTCATGCGGCTCT	1133
QY	901	ACCCACCGTATGCGGGGGTGGTGGATTCCTGTACTCCGCGCCCTTGCTTGAGCGTGT	960
Db	1134	ACCCACCGTATGAGGGGGAGGGGGTTCCTTACTCCGCGCCACCTGGCGCTGAGCGTGT	1193
QY	961	ACAGTCGACTAGCCGGTCCATCTCTACCTTATGATGATGATGATGATGATGATGATGATG	1020
Db	1194	ACCATATCTAGTACGAGGTCATCTCTACCTTATGATGATGATGATGATGATGATGATGATG	1253
QY	1021	TTCAAGAACTGGGCTTGTTCAGAGAGCAAGCAAGGCTTCAGGACATTTGATATTGAAG	1080
Db	1254	TTCAAGAACTGGGCTTGTTCAGAGAGCAAGCAAGGCTTCAGGACATTTGATATTGAAG	1313
QY	1081	AGAAAATAGAAAATATTTTGTCTATATAGACCTAATGTTAGTACATAGCAGAAAC	1140
Db	1314	AGAAAATAGAAAATATTTTGTCTATATAGACCTAATGTTAGTACATAGTACAGAAAC	1373
QY	1141	CTCAAGAGATGATTGATATCTGGTCTCAGTTGCAAGGCTTAATTTAAAGTCTGA	1196
Db	1374	CTCAAGAGATGATTGATATTGGTCTCAGTTGCAAGGCTTAATTTAAAGTCTGA	1429
RESULT 8			
LOCUS	BD093191	1912 bp	DNA linear PAT 27-AUG-2002
DEFINITION	Useful polypeptide.		
ACCESSION	BD093191		
VERSION	BD093191.1 GI:22638779		
KEYWORDS	WO 0100848-A/1.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 1912)		
	Sasaki, K., Shiraishi, N., Natsume, A., Yamada, Y., Nakagawa, S. and		
	Sekine, S.		
TITLE	Useful polypeptide		
JOURNAL	Patent: WO 0100848-A 1 04-JAN-2001;		
	KYOWA HAKKO KOGYO CO LTD, KATSUTOSHI SASAKI, NORIHIKO SHIRAISHI,		
	AYUMI NATSUME, YOJI YAMADA, SATOSHI NAKAGAWA, SUSUMU SEKINE		
COMMENT	OS Homo sapiens (human)		
	PN WO 0100848-A/1		
	PD 04-JAN-2001		
	PE 29-JUN-2000 WO 2000JP004304		
	PR 29-JUN-1999 JP 99p 183437, 16-MAR-2000 JP OCP 074757 PI		
	KATSUTOSHI SASAKI, NORIHIKO SHIRAISHI, AYUMI		
	NATSUME, YOJI YAMADA,		
	PI SATOSHI NAKAGAWA, SUSUMU SEKINE		
	PC C12N15/54, C12N9/10, C12N5/10, C12N1/21, A61K48/00, A61K45/00, PC		
	A61K39/395,		
	PC A61K35/00, A61K31/71, A61P35/00, A61P29/00, A01K67/027, A01H5/00,		
	PC G01N33/53,		
	PC G01N33/15, G01N33/05		
	CC		
FEATURES	Key Location/Qualifiers.		
source	Location/Qualifiers		
	1..1912		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
BASE COUNT	518 a 430 c 457 g 507 t		
ORIGIN			
Query Match	75.8%; Score 906.4; DB 6; Length 1912;		
Best Local Similarity	84.9%; Pred. No. 1.5e-244;		
Matches 1015; Conservative	0; Mismatches 181; Indels 0; Gaps 0;		

QY	1	AGATGAGTGTGGCGTCGAAGAGTCAAGTTGCTGGCATCTCTGATGATGCGCAATGTCT	60
Db	232	AAATGAGTGTGGACGTGGAAGATAAAGTTGTGGGTATCTTGATGATGCGCAATGTCT	291
QY	61	TCATTTATTTGATGTGGAAGTCTCCAAAACAGTAGCCAAAGACAAAATGGAAGGGAG	120
Db	292	TCATTTATTTTATTTATGGAAGTCTCCAAAAGCAGTAGCCAAAGAAAATGGAAGGGG	351
QY	121	GAGTAATAATCCCAAGAGAGAAAGTTCTGGAAGCCACCCAGCACTCCCGGGCATACTGGA	180
Db	352	AAGTAATAATACCAAGAGAGAGTTCTGGAAGATATCTACCCCTCCCGAGGCATAGGA	411
QY	181	ACAGGGAACAGAGAGAGCTGAACAGTGGTACAAATCCATCTTTGAACAGSGTGGCCAACT	240
Db	412	ACCAGAGCAAGAGAGAGCTGAACCGGAGTACAAACCCCATCTCTGAGCATCTGACCAACC	471
QY	241	AGACAGGGGAGTACGCACATCTCCAAACACAAAGTCACTGAGCTATTGTGAACACAGACT	300
Db	472	AGACGGGGAGCGCGGAGGCTCTCCAAATATAGCCATCTGAACACTACTGCAACCTGACC	531
QY	301	CGACGGTCATGACAGCTGTGACAGATTTTAAATATCTGCCGGACAGATTTTAAAGACTTTC	360
Db	532	TGAGGGTCACGTCGGTGGTTACGGGTTTAAACAACTTTCGGGACAGATTTTAAAGACTTTC	591
QY	361	TCCTGTATTTGAGATGCGGGAATTAATCGTGTCTATAGATCAACGCAAGAAATGTGCA	420
Db	592	TGCTGTATTTGAGATGCGGCAATTAATCACTGCTTATAGATCAACGCGGATAGGTGCAA	651
QY	421	AGAAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATCTCCACATTTTGCAGAGCAAG	480
Db	652	AGAAACCTTCTTCTGTGTCGGGATTAAAGTCCCTCACTCCACATTTTGCAGAGCAAG	711
QY	481	CAATTCGGGAGTCTTGGGGCGGAGAAACCAACAGTAGGAAACCAAGTAGTACAGGTCT	540
Db	712	CAATTCGGGGAATCTCGGGGCCAAGAAAGCAACGAGGGAACCAAGCGTGGTGGAGTCT	771
QY	541	TCCTGTTGGGCAAGACACACCCCGAGGACCAACCCCTGACCTTTCGGACATGCTTAAAGT	600
Db	772	TCCTGCTGGCGAGACACACCCCGAGGACCAACCAACCCCGACCTTTCAGATATGCTGAAT	831
QY	601	TTGAGAGTGAACAGACACACAGGACATCTCTATGTGGAACACTATAGAGACACATTTCTCAACC	660
Db	832	TTGAGAGTGAAGACACACAGGACATCTTATGTGGAACACTACAGAGACACTTTCTCAACT	891
QY	661	TGCTCCTGAAGAAAGTCTGTTCTTAGTGGGTGAGCACTTCTGTCAGAGACGAGAGT	720
Db	892	TGCTCCTGAAGAAAGTCTGTTCTCAGTGGGTGAAGTACTTCTGCGCCAGACACTGAGT	951
QY	721	TTGCTTCAAGGGGATGATGACGCTGTTTGTGAACACCCATCATCTTAAATACTTGA	780
Db	952	TTGCTTCAAGGGGATGACGATGTTTGTGAACACCCATCATCTTAAATACTTGA	1011
QY	781	ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTTCATAGGTGAGCTGATCCCAATGCTG	840
Db	1012	ATAGCTTATCCAGAGCAAGCCAAAGACTTCTTTCATAGGTGAGCTGATCCCAATGCTG	1071
QY	841	GGCTCACCAGGTAAGCAAACTGAAGTACTACATCCCGAGAGTCTTCTACACCGGGGTCT	900
Db	1072	GACCTATCGGTAAGCAAACTGAAGTACTACATCCCGAGAGTCTTCTACTCTGGCTCT	1131
QY	901	ACCCACCGTATGCGGGGGTGGTGGATTCCTGTACTCCGGCCCTTGCCCTTGAGGCTGT	960
Db	1132	ACCCACCGTATGAGGGGGAGGGGGTTCCTTACTCCGGCCACCTGGCCCTGAGGCTGT	1191
QY	961	ACAGTGGCACTAGCGGTCATCTCTACCTATTTGATGATGTTTATACGGGATGTGCC	1020
Db	1192	ACCATATCTGACCAAGTCCATCTCTACCCCATTTGATGAGCTTTATCTGGAATGTGCC	1251
QY	1021	TTCAAGAACTGGGCTTGTTCAGAGAGCAAGGCTTCAGGACATTTGATATTGAAG	1080
Db	1252	TTCAAGAACTGGGCTTGTTCAGAGAGCAAGGCTTCAGGACATTTGATATTGAAG	1311


```

RESULT 10
LOCUS BC030579
DEFINITION Homo sapiens, UDP-GlcNAc:betaGal
beta-1,3-N-acetylglucosaminyltransferase 1, clone MGC:26071
IMAGE:4828158, mRNA, complete cds.
ACCESSION BC030579
VERSION 1
KEYWORDS GI:21040508
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2454)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 34 Row: d Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 15451893.
Location/Qualifiers
1. .2454
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:10678"
/db_xref="taxon:9606"
/clone="MGC:26071 IMAGE:4828158"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/notes="Vector: pBluescript"
113..1306
/codon_start=1
/product="UDP-GlcNAc:betaGal
beta-1,3-N-acetylglucosaminyltransferase 1"
/protein_id="AAH30579.1"
/db_xref="GI:21040508"
/translation="MSVGRRIKKLGLILMANVFIYFIMEVSKSSSEKNGKEVITP
KEKFWKTPPEAYNROEKLNRQYNPILSMLNOTGEAGRLSNISHLNYCEPLRV
TSVATFTNLPDRKDFLLYLRCRNYSLIDQPKCAKKPELLLAISLTPHFAARQA
IRESQGSNAGNQPVRVFLGQTPPEDNHPDLSMDLKFSEKHODILMWNVYRDTF
NLSLKEVLFLRWYSQCPDTEPVFKGDDDDVFNTHILNLSLTKFKAKDLFIQDVI
HNAGPHRQKKLYIVVYSGLYPPYAGGGGFLYSGLHRLRLRHITDQVHLYPIDV
YTGMLQKLGIVPEKHGKGFRTDIEKNKNICSYVDLMLVRSRKPQEMIDISQLOS
AHLK"
BASE COUNT 725 a 444 c 524 g 761 t
ORIGIN
Query Match 75.8%; Score 906.4; DB 9; Length 2454;
Best Local Similarity 84.9%; Pred. No. 1.6e-244;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
QY 1 AGATGAGTGTGGCGCTCGAAGAGTCAAGTTCGICGGGATCCTGATGTCGCAATGCTC 60
111 AAATGAGTGTGGCGCTCGAAGATAAAGTTGTGGTATCCTGATGTCGCAATGCTC 170
61 TCATTTATTTGATTTGGGAAGTCTCCAAAACAGTAGCAAGCAAAATATGGAAGGAG 120
171 TCATTTATTTTATTTATGGAAGTCTCCAAAAGCAGTAGCAAGCAAAATATGGAAGG 230
121 GAGTAATAATCCGAAAGAGAGTCTCGAAGCCAGCCAGCACTCCCGGGGATACATGA 180
231 AAGTAATAATACCCAAGAGAGTCTCGAAGATATCTACCCCTCCCGAGGATATCTGA 290
181 ACAGGAACAGGAGAGAGTGAACAGGTGTAGCAATCCCATCTTGAACAGGGTGGCAATC 240
291 ACCGAGAGCAAGAGAGAGTGAACCGGCACTACAACCCCATCTGAGCATGCTGACCA 350
241 AGACAGGGAGTGTAGCCACATCTCCAACACAGTCACTGAGCTATTGTGAAACGACT 300
351 AGACGGGGAGGCGGCGAGGCTCTCCAATATAGCCATCTGAACCTACTGCAACCTGAC 410
301 CGACGGTCAATGAGCTGTGACAGATTTTAATAATCTGCCGACAGATTTAAGACTTTC 360
411 TGAGGTCACTGCTGGTGTACGGGTTTAAACAATCTGCCGACAGATTTAAGACTTTC 470
361 TCTTGTATTGAGATCCCGGAATTACTCGCTGCTTATAGATCAACCGAAGAAATGTGCA 420
471 TGCTGTATTGAGATCCCGAATTATTACTGCTTATAGATCAGCGGATAGTGTGCAA 530
421 AGAAGCCCTTCTACTATTGGCGATAAAGTCCCTCATTCACATTTGCCAGAGCAAG 480
531 AGAACTTTCTTTGTTGCTGGGATTAAGTCCCTCACTCCACATTTGCCAGAGCAAG 590
481 CAATCCGGAGTCTTGGGGCCGAGAAACCAAGTAGGGAACAGACAGTAGTAGGGTCT 540
591 CAATCCGGGAATCTTGGGCCCAAGAACCAAGCAGGAGAACCAAGGCTGTGCGAGTCT 650
541 TCCTGTGGCAGACACACCCAGAGAGCAACCACTGAGCTTTTGGACATGCTTAAGT 600
651 TCCTGTGGGCCAGACACCCAGAGAGCAACCACTGAGCTTTTGGACATGCTTAAGT 710
601 TTGAGAGTGAACAAGCACCAGGACATCTCTGTGGAACTATAGAGACACATCTTCAAC 660
711 TTGAGAGTGAAGAAGCACCAGACATCTTATGTGGAACATACAGACACATCTTCAACT 770
661 TGTCCTTGAAGGAGTGTGTTTCTTAGTGGGTGAGCACTTCTCTGTCAGACGAGAGT 720
771 TGTCCTTGAAGGAGTGTGTTTCTTAGTGGGTGAGCACTTCTCTGTCAGACGAGT 830
721 TGTCCTTGAAGGCGATGATGAGTGTGTTGTAACACCCATCACATCTTAAATTCGA 780
831 TTGTTTTCAGGCGGATGAGGATGTTTGTGTAACACCCATCACATCTTAAATTCGA 890
781 ATAGCTTATCCAGAGCAAGCAAGAGTGTGTTTATAGGTGAGCTTATCCACAAATGCTG 840
891 ATAGTTATCCAGAGCAAGCAAGAGTGTGTTTATAGGTGAGCTTATCCACAAATGCTG 950
841 GGCTCACCAGGATAGAAACAGTACTACATCCCAAGAGTCTTACACACGGGCTCT 900
951 GACCTCATCGGATAGAAAGTGAAGTACTACATCCCAAGAGTGTGTTTACTCTGCGCTCT 1010
901 ACCACCGTATCGCGGGGTGGTGGATTCCTGTACTCCCGGCCCCCTTGCTTGGAGGCTGT 960
1011 ACCACCGTATCGCGGGGTGGTGGATTCCTGTACTCCCGGCCCCCTTGCTTGGAGGCTGT 1070
961 ACATGCGCATAGCGGGGTCCCATCTCTACCCCTATTGATGATGTTTATACGGGAATGTGC 1020
1071 ACCATATCATGACAGGTCCTCTACCCCATGATGACGTTTATATCGGATGTGCC 1130
1021 TCCAGAACTGGCCCTTGTTCAGAGAGCAACAAAGCTTTCAGGACATTTGATATTTGAAG 1080
1131 TCCAGAACTGGCCCTTGTTCAGAGAGCAACAAAGCTTTCAGGACATTTGATATTCGAGG 1190
1081 AGAAATATAGAAATATTTGTTCCCTATATAGACCTTAATGTAGTACATACAGAAAC 1140
1141 AGAAATATAGAAATATTTGTTCCCTATATAGACCTTAATGTAGTACATACAGAAAC 1190

```

Db	1191	AGAAAACAAAATACATCTGCTCTCTATGTAGATCTGATGTAGTAGACATAGTAGAAAC	1250
Qy	1141	CTCAAGAGATGATTGATATCTGCTCTCAGTTGCAGAACTCTCAATTTAAAAATGCTGA	1196
Db	1251	CTCAAGAGATGATTGATATTTGGTCTCAGTTGCAGAGTGTCTATTTAAAAATGCTAA	1306
RESULT 11			
AF288208			
LOCUS	AF288208	2714 bp mRNA linear	PRI 03-AUG-2000
DEFINITION	Homo sapiens galactosyltransferase beta3Gal-76 mRNA, complete cds.		
ACCESSION	AF288208		
VERSION	AF288208.1	GI:9664886	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 2714)		
TITLE	Gromova,I., Gromov,P. and Celis,J.		
JOURNAL	A novel member of beta-1,3-galactosyltransferase family is down		
REFERENCE	regulated during bladder ICC progression		
AUTHORS	unpublished		
TITLE	2 (bases 1 to 2714)		
JOURNAL	Gromova,I., Gromov,P. and Celis,J.		
FEATURES	Direct Submission		
SOURCE	Submitted (18-JUL-2000); Medical Biochemistry, University of Aarhus.		
LOCUS	Ole Worms Alle 170, Aarhus C-8000, Denmark		
DEFINITION	Location/Qualifiers		
LOCUS	1..2714		
DEFINITION	/organism="Homo sapiens"		
LOCUS	/mol_type="mRNA"		
DEFINITION	/db_xref="taxon:9606"		
LOCUS	/tissue_type="bladder"		
DEFINITION	229..1422		
LOCUS	/codon_start=1		
DEFINITION	/product="galactosyltransferase beta3Gal-76"		
LOCUS	/protein_id="AAF97253.1"		
DEFINITION	/db_xref="GI:9664887"		
LOCUS	/translation="MSVGRRIKLLGILMANVFIYFIMEVSKSSSQEKNKGVEILIP		
DEFINITION	KEFKWISITPPEAYWNRDEKLNQNPILSLTNOTGEGRLNSINSHLNYCEPDLRV		
LOCUS	TSVYTFNNLPDRFKDLFLRCRNYSLIIDQDKCAKFPFLLLAKSLTTPHARQVA		
DEFINITION	IRSEGWESNAGNQTVPVRFLLGTPEDDHPDLSMLKPESEKHQDILMNYRDYFF		
LOCUS	NLSLKEFLWYSTQCTPEFYKGDNDVNTLHILNLYNSLSTKAKDLFGDVI		
DEFINITION	HNAGPRDKKLKYYIEVYSGLYPPYAGGGFLYSGHLALRLYHITDOVHLYPDIV		
LOCUS	YTCMCKLKLGLVPEKHKGRTFDIEEKNKNICSYVDLMLVHRSKQEMDIIWSQLS		
DEFINITION	AHLKC"		
BASE COUNT	777 a 515 c 601 g 821 t		
ORIGIN			
Query Match	75.8%	Score 906.4;	DB 9; Length 2714;
Best Local Similarity	84.9%;	pred. No. 1.6e-244;	
Matches 1015;	Conservative	0; Mismatches 181;	Indels 0; Gaps 0;
Qy	1	AGATGAGTGTGGGGGTCGGAAGAGTCAAGTTGCTGGGCATCCTGATGATGGCAATGTCT	60
Db	227	AAATGAGTGTGGACGTCGAAGATAAAGTGTGTGGGTATCCTGATGATGGCAATGTCT	286
Qy	61	TCATTTATTTGATTGTGGAGTCTCCAAAACAGTAGCCACAGACAAAATGGAAAGGAG	120
Db	287	TCATTTATTTATTTATGGAAGTCTCCAAAAGCAGTAGCCAAAGAAAATGGAAAGGGG	346
Qy	121	GAGTAATATCCGAAAGAGAAGTTCTGGAAGCCACCCAGCAGTCCCGGGCATCTGGA	180
Db	347	AAGTAATAATCCCAAGAGAAGTCTTGGAAGATAATCTACCCCTCCCGAGGCATCTGGA	406
Qy	181	ACAGGGAAGAGAGAAGCTGAACAGGTGGTGAACATCCCATCTTGAACAGGGTGGCCAAATC	240
Db	407	ACCGAGAGCAAGAGAAGCTGAACGGGAGTAGACACCCCATCTCTGAGCATCTGACCAACC	466
Qy	241	AGACAGGGGAGCTAGCCACATCTCCAACACAGTCACTGAGCTATTGTGAACACAGACT	300
Db	467	AGACGGGGAGGGCGGCGGCTCTCCAATATAAGCCATCTGAACTGCTGGAACCTGACCC	526

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Hominidae: Homo. 1 (bases 1 to 2811) Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Blatt,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,A.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raja,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullah,J.S., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,L.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Maman,A., Rodriguez,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.E., Schnerch,A., Schein,J.E., Jones,S.J., and Marra,M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002) 22389257 12477932 2 (bases 1 to 2811) Strausberg,R. Direct Submission Submitted (03-MAR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA NIH-MGC project URL: <http://mgc.nci.nih.gov> Contact: MGC help desk Email: cgaps-femail.nih.gov Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre. BC Cancer Agency, Vancouver, BC, Canada info@bcgsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjeil, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeeadi, Jacqueline Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov> Series: IRAC Plate: 89 Row: 3 Column: 21 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 15451893. FEATURES source 1. .2811 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:49838 IMAGE:5767816" /tissue_type="Brain, fetal, whole pooled" /clone_lib="NIH_MGC_L21" /lab_host="DH10B" /note="vector: pCMV-SPORT6" 1. .2811 /gene="B3GNT1" /note="synonyms: BETA3GNT, B3GNT-2, B3GN-T2, B3GN-T1, B3GN1" /db_xref="LocusID:10678" /db_xref="MIM:605581" 239. .1432 CDS

/codon_start=1 /product="beta-1,3-N-acetylglucosaminyltransferase bGnt-1" /protein_id="AAH47933.1" /db_xref="GI:28839675" /db_xref="LocusID:10678" /translation="MSVRRRTKLGLIMANVFIYFIMEVSKSSQKNGKEVILP KEKFWKISTPTPEAYWNRQEKLRQNYNPILSDNOTGAGRLSNISHLNYCEPDLRV TSVVTGNLPLDRFKDFLLVLCRNYSLLIDQPKCAKPFLLAYKSLTPHARRQA IRESWGOESNAGNQTVYRVFLLQOTPPDNHPLDLSMLAFSEKSHODILMWNRYRTFF NLSLKEVFLRWVSTPCDFEYVFKGDDVFNTHILNLFNLSLSTKRDQLPIGDVI HNSGPHRDKLKYIIPVYVSYGLPPYAGGGGLYSGHLALRLYHITDQVHLPIDDV YHLCMLQKLGVLPEKHKHFTFDIEEKNKNNICSYVDLMLVSHRSPQEMIDIWSQLOS AHLKC" BASE COUNT 837 a 527 c 613 g 834 t ORIGIN Query Match 75.8%; Score 906.4; DB 9; Length 2811; Best Local Similarity 84.9%; Pred. No. 1.6e-244; Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0; QY 1 AGATGAGTGTGGGCGCTCGAAGACTCAAGTTGCTGGGCATCCTGATGATGCGAAATGTCT 60 DB 237 AAATGAGTGTGGGCGCTCGAAGATAAAGTTGTTGGGTATCTCGATGATGCGAAATGTCT 296 QY 61 TCATTTATTTGATTGTGGAAGTCTCCAAAACAGTAGCCAAAGACAAAATGGAAGGGAG 120 DB 297 TCATTTATTTATTTATGGAAGTCTCCAAAAGCAGTAGCCAAAGAAAATGGAAGGGG 356 QY 121 GAGTAATAATCCCGAAAGAGAGTTCTTGGAGCCGCCAGCACTCCCGGGGCATCTGGA 180 DB 357 AAGTAATAATACCGAAAGAGAGTTCTTGGAGATATCTACCCCTCCCGAGGACTACTGA 416 QY 181 ACAGGGAACAGGAGAAGCTGAACAGGTGTACATCCATCTTGAACAGCGGTGGCAATC 240 DB 417 ACCGAGCAAGAGAGAGCTGAACCGGCGAGTACAAACCCATCTCGAGCATGCTGACCA 476 QY 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAAAGTCACTGAGCTATTTGGAACCAAG 300 DB 477 AGACGGGGAGGCGGCGAGTCTCCCAATATAGCCATCTGAACTACTGCAACCTGACC 536 QY 301 CGACGGTCATGACAGCTGTGACAGATTTTAAATATCTGCCGGACAGATTTAAAGACTT 360 DB 537 TGAGGGTCACGTCGGTGTTCACGGGTTTAAACAACTTGCCGGACAGATTTAAAGACTT 596 QY 361 TCTTGTATTTCAGATGCGGGAATTTACTCGCTGCTTATAGATCAACCCGAAGAAATGTG 420 DB 597 TGCTGTATTTCAGATGCGGCAATTTACTCACTGCTTATAGATCAGCGGAGATAGTGTG 656 QY 421 AGAAGCCCTCTTACTATTGCGGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 480 DB 657 AGAAACCTTCTTGTGTGCGGATTAAGTCCCTCACTCCACATTTTCCAGAGGCAAG 716 QY 481 CAATTCGGGAGTCTTGGGGCGGAGAACCAACAGTAGGGAACACAGACAGTAGTGGAGT 540 DB 717 CAATCCGGGATCTTGGGGCCAGAGAGACGAGGAGGAGGAGGAGGAGGAGTGTGCGAG 776 QY 541 TCCTGTTGGGCAAGACACCCCGGAGGAGACACACCTGACCTTTCGGACATGCTTAAGT 600 DB 777 TCCTGTTGGGCGAGACACCCCGGAGGAGACACACCCCGGAGCTTTCAGATATGCTGAA 836 QY 601 TTGAGAGTGACAAGCACACGAGACATCTCATGTGGAATATAGAGACACATCTTCAACC 660 DB 837 TTGAGAGTGAGAACACACGAGACATCTTATGTGGAATATAGAGACATCTTCAACT 896 QY 661 TGTCCCTGAAGAAAGTGTCTTTCTTAGTGGGTGAGCACTTCTGTCAGAGAGGAGAGT 720 DB 897 TGTCTGGAAGAAAGTGTCTTTCTCAGTGGGTAACTACTTCTCGCCAGACACTGAGT 956 QY 721 TTGCTCTCAAGGGGAGTGTGAGCTGTTTGTGAACACCCATCATCTTAAIATCTTGA 780 DB 957 TTGTTTCAAGGGGAGTGTGAGCTGTTTGTGAACACCCATCATCTTAAIATCTTGA 1016 QY 781 ATAGCTTATCCAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCTG 840


```
misc_feature 3789..4320 /note="match to EST BG721682 (NID:g14000869)"
misc_feature 3829..4087 /note="match to EST AA377082 (NID:g2029630)"
misc_feature 3869..4691 /note="match to EST AU137528 (NID:g10998067)"
misc_feature 4109..4563 /note="match to EST H78875 (NID:g1056964) yu09g06.rl"
misc_feature 4141..4441 /note="match to EST AA263148 (NID:g:898954)"
misc_feature 4142..4793 /note="match to EST BG741357 (NID:g:4052010)"
misc_feature 4180..4678 /note="match to Mus musculus EST AA204363 (NID:g1801007) mu30ell1.rl"
misc_feature 4264..4724 /note="match to EST AW958548 (NID:g8148232)"
misc_feature 4426..5325 /note="similar to EST BM557603 (NID:g18799721)"
misc_feature 4468..4938 /note="similar to Mus musculus EST BE632125 (NID:g9914895) uu57d09.yl"
misc_feature 4498..4879 /note="match to EST H47990 (NID:g924042) yf75d11.rl"
misc_feature 4624..5231 /note="match to EST BF001654 (NID:g10701929)"
misc_feature 4666..5089 /note="match to EST AV682596 (NID:g10284459)"
misc_feature 4680..5391 /note="match to EST BE622082 (NID:g98993022)"
misc_feature 4915..5257 /note="match to EST R31722 (NID:g787565) yb63d09.rl"
misc_feature 4945..5390 /note="match to EST H13125 (NID:g877945) yj06e09.rl"
misc_feature 4948..5597 /note="match to EST BF131058 (NID:g10970087)"
misc_feature 5026 /note="match to EST R74552 (NID:g848922) y158a02.rl"
misc_feature 5031..5682 /note="match to EST BE564288 (NID:g9808008)"
misc_feature 5061..5553 /note="match to EST N58174 (NID:g1202064) yv66a07.sl"
misc_feature 5095..5554 /note="match to EST N66915 (NID:g1219040) za47c09.sl"
misc_feature 5107..5599 /note="match to EST R74552 (NID:g848922) y158a02.rl"
misc_feature 5128..5793 /note="match to EST BE865533 (NID:g10314309)"
misc_feature 5220..5836 /note="match to EST BF132275 (NID:g10971315)"
misc_feature 5254..5534 /note="match to EST BG319463 (NID:g13129329)"
misc_feature 5289..6180 /note="similar to EST BM550513 (NID:g18786776)"
misc_feature 5302..5473 /note="similar to Bos taurus EST BM087084 (NID:g16997712)"
misc_feature 5316..5838 /note="match to EST AA773646 (NID:g7825217) af75c02.rl"
misc_feature 5359..5986 /note="match to EST AV724225 (NID:g10828399)"
misc_feature 5482..6241 /note="match to EST BE738519 (NID:g10152511)"
misc_feature 5483..6278 /note="similar to Homo sapiens EST B5622910 (NID:g9893850)"
misc_feature 5503..5885 /note="similar to Bos taurus EST BM087084 (NID:g16997712)"
misc_feature 5641..6134 /note="match to EST AV715275 (NID:g10796792)"
misc_feature 5648..5885 /note="similar to Sus scrofa EST BF075149 (NID:g10868660)"
misc_feature 5682..6177 /note="match to EST AV705681 (NID:g10722976)"
```

```
misc_feature 5802..6310 /note="match to EST BF109367 (NID:g10939057)"
misc_feature 6009..6312 /note="match to EST AA835126 (NID:g2908854) ak63e07.sl"
misc_feature 6009..6311 /note="match to EST AV659072 (NID:g9880086)"
repeat_region 6010..6038 /rpt_family="AI-rich"
misc_feature 6224..8480 /note="match to EST D87736 (NID:g1552215)"
misc_feature 6229..6310 /note="match to EST D87736 (NID:g1552215)"

O-ery Match 75.8%; Score 906.4; DB 9; Length 99088;
Best Local Similarity 84.9%; Pred. No. 3.1e-244;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGCGTCGAAGAGTCAAGTTCGTCGGCATCCTGATGATGCGCAATGTCI 60
Db 3800 AAATGAGTGTGGGCGTCGAAGAAATAAGTTGTTGGGTATCCTGATGATGCGCAATGTCI 3859
QY 61 TCATTATTGATGTGGAAGTCTCCAAAACAGTAGCCAGACAAAATGGAAGGGAG 120
Db 3860 TCATTATTATTATGGAAGTCTCCAAAACAGTAGCCAGACAAAATGGAAGGGAG 3919
QY 121 GAGTAATAATCCGAAAGAGAGTCTGGAAGCCACCAGCACCTCCCGGGGCATACCTGGA 180
Db 3920 AAGTAATAATCCGAAAGAGAGTCTGGAAGATATCTACCCCTCCCGGGGCATACCTGGA 3979
QY 181 ACAGGGAACAGAGAGCTGAACAGGTGGTACATGCCATCTTGAACAGGGTGGCCAACTC 240
Db 3980 ACCGAGAGCAAGAGAGCTGAACAGGTGGTACATGCCATCTTGAACAGGGTGGCCAACTC 4039
QY 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACACAGCT 300
Db 4040 AGACAGGGGAGCGGGCAGGCTCTCCAAATATGAGCATCTGAACTACTGCAACCTGACC 4099
QY 301 CGACGGTCATGACAGCTGTGACAGATTTTAAATAATCTGCGGAGACAGATTTAAAGACTTC 360
Db 4100 TGAGGTCACGTCGGTGGTTACGGGTTTTAAACAACTTCCCGGACAGATTTAAAGACTTC 4159
QY 361 TCTTGTATTGAGATGCGGGAATTAATCCTGCTTATAGATCAACCGAAGAAATGTCGA 420
Db 4160 TCTGTATTGAGATGCGGGAATTAATCCTGCTTATAGATCAACCGGGAATGTCGA 4219
QY 421 AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCATTCACATTTGCCAGAGAGCAAG 480
Db 4220 AGAAGCCCTTCTTGTGTCGCGATTAAAGTCCCTCATTCACATTTGCCAGAGAGCAAG 4279
QY 481 CAATTGGGAGTCTTGGGCGGAGAAACCAACGATAGGGAACACAGACAGTAGTAGGTCI 540
Db 4280 CAATCGGGAATCTTGGGCGGAGAAACCAACGATAGGGAACCAACGATAGGTCI 4339
QY 541 TCTGTGGGCAAGACACCCCGAGAGCAACCCCTGACCTTTCGGACATGCTTAAGT 600
Db 4340 TCTGTGGGCAAGACACCCCGAGAGCAACCCCTGACCTTTCGGACATGCTTAAGT 4399
QY 601 TTGAGAGTGAACAGCAGGACATCTTCATGCTGGAATATAGAGACATCTTCAACC 660
Db 4400 TTGAGAGTGAACAGCAGGACATCTTCATGCTGGAATATAGAGACATCTTCAACC 4459
QY 661 TGTCTCTGAAGGAAGTGTCTTCTTAGTGTGGTGAGCACTTCTGTCACAGCAGAGT 720
Db 4460 TGTCTCTGAAGGAAGTGTCTTCTTAGTGTGGTGAGCACTTCTGTCACAGCAGAGT 4519
QY 721 TTGTCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCATCTTAAATTAATGTA 780
Db 4520 TTGTCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCATCTTAAATTAATGTA 4579
QY 781 ATAGCTTATCCAAAGAGCAAGCAAGACTTGTTCATAGGTGAGCTGATCCACAAATGCTG 840
Db 4580 ATAGCTTATCCAAAGAGCAAGCAAGACTTGTTCATAGGTGAGCTGATCCACAAATGCTG 4639
QY 841 GGCCTCACCGGATGAAGAACTGAAGTACTATACCCAGAGAGTCTTCTACACCGCGTCT 900
```


PA (MILL-) MILLENNIUM PHARM INC.
 XX White D, Zhou J, Tartaglia LA;
 PI WPI: 2000-271461/23.
 XX P-PSDB; AAY84641.
 XX
 PT Method for determining compounds which modulate body weight and can be
 PI used to treat e.g. obesity comprises measuring the activity of leptin
 PT inducible genes -
 XX
 XX Claim 16: Fig 1A-B: 123pp: English.
 PS
 XX The present sequence encodes a leptin induced LIG46 polypeptide. The
 CC specification describes a method for determining whether a compound can
 CC be used to modulate body weight by measuring the activity of leptin
 CC inducible genes, such as LIG46. The method can be used to specifically
 CC detect LIG46 and LIG56 nucleic acid molecules relative to other nucleic
 CC acid molecules encoding galactosyltransferases or GTP-binding proteins.
 CC They can also be used in diagnostic assays to identify the presence or
 CC absence of a genetic lesion or mutation characterized by aberrant
 CC modification or mis-regulation of the genes or aberrant post-
 CC translational modification of the proteins. LIG46 and LIG56 proteins
 CC and nucleic acid molecules can be used to treat obesity or cachexia.
 CC LIG46 and LIG56 antagonists are used to treat obesity and LIG46 and
 CC LIG56 agonists are used to treat low body weight. The leptin inducible
 CC genes i.e. LIG46, LIG56, Tgtp, LRG-47, RC10-II and Stral3 can be used
 CC to determine whether a compound modulates body weight and can then be
 CC used to treat obesity or cachexia or low body weight.
 XX
 XX Sequence 1196 BP: 350 A; 269 C; 290 G; 287 T; 0 other;
 SQ
 Query Match 100.0%; Score 1196; DB 21; Length 1196;
 Best Local Similarity 100.0%; Pred No. 0;
 Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGCGTCGGAAGAGTCAAGTGTCTGGGCATCTCTGATGATGGCAATGTCT 60
 DB 1 AGATGAGTGTGGGCGTCGGAAGAGTCAAGTGTCTGGGCATCTCTGATGATGGCAATGTCT 60
 QY 61 TCATTTATTTGATTTGGAGTCTCCAAACACAGTAGCCAGACAAAATGGAAGGGAG 120
 DB 61 TCATTTATTTGATTTGGAGTCTCCAAACACAGTAGCCAGACAAAATGGAAGGGAG 120
 QY 121 GAGTAATAATCCCGAAGAGAGTCTTGGAGGCCACCCAGCAGCTCCCGGGCATACTGGA 180
 DB 121 GAGTAATAATCCCGAAGAGAGTCTTGGAGGCCACCCAGCAGCTCCCGGGCATACTGGA 180
 QY 181 ACAGGGACAGAGAGAGCTGAACAGGTGTGTAATCCCATCTTGAACAGGGTGGCCATC 240
 DB 181 ACAGGGACAGAGAGAGCTGAACAGGTGTGTAATCCCATCTTGAACAGGGTGGCCATC 240
 QY 241 AGACAGGGAGCTAGGCACATCTCCAAACACAGTCACCTGACCTATCTGACACAGCT 300
 DB 241 AGACAGGGAGCTAGGCACATCTCCAAACACAGTCACCTGACCTATCTGACACAGCT 300
 QY 301 CGACGGTCATGACAGCTGTGACAGATTTTAATATCTGCGGACAGATTTAAGAGCTTC 360
 DB 301 CGACGGTCATGACAGCTGTGACAGATTTTAATATCTGCGGACAGATTTAAGAGCTTC 360
 QY 361 TCTGTATTGAGATCCCGAANTACTCGTGTCTTAGATCAACCGAAGAAATGTGCA 420
 DB 361 TCTGTATTGAGATCCCGAANTACTCGTGTCTTAGATCAACCGAAGAAATGTGCA 420
 QY 421 AGAAGCCCTCTTACTATTGGCGATAAGTCCCTCATCTCCACATTTGGCAGAGGCAAG 480
 DB 421 AGAAGCCCTCTTACTATTGGCGATAAGTCCCTCATCTCCACATTTGGCAGAGGCAAG 480
 QY 481 CAATTCGGAGTCTTGGGCGCGAAGAACCAACAGCTAGGGAACACAGAGTAGTGAGGCTCT 540
 DB 481 CAATTCGGAGTCTTGGGCGCGAAGAACCAACAGCTAGGGAACACAGAGTAGTGAGGCTCT 540
 QY 541 TCCTGTTGGGCAAGACACCCCGAGAGGACAAACCCCTGACCTTTCCGACATGCTTAAGT 600

DB 541 TCCTGTTGGGCAAGACACCCCGAGAGACAAACCCCTGACCTTTCCGACATGCTTAAGT 600
 QY 601 TTGAGAGTGAACAGCACCAGGAGACATCTCTATGTGGAACTATAGAGACACATTTTCAACC 660
 DB 601 TTGAGAGTGAACAGCACCAGGAGACATCTCTATGTGGAACTATAGAGACACATTTTCAACC 660
 QY 661 TGTCCTTGAAGGAGTGTCTGTTTCTAGGTGGTGAGCAGCTTCTCTCCAGAGCGAGAGT 720
 DB 661 TGTCCTTGAAGGAGTGTCTGTTTCTAGGTGGTGAGCAGCTTCTCTCCAGAGCGAGAGT 720
 QY 721 TTGCTTCAAGGGGATGATGAGTGTCTGTGAACACCCCATCATCTCTTAATTAATTCTGA 780
 DB 721 TTGCTTCAAGGGGATGATGAGTGTCTGTGAACACCCCATCATCTCTTAATTAATTCTGA 780
 QY 781 ATAGCTTATCCAAAGAGCAAAAGCAAGCTTGTTCATAGGTGACGTGATCCCAATGCTG 840
 DB 781 ATAGCTTATCCAAAGAGCAAAAGCAAGCTTGTTCATAGGTGACGTGATCCCAATGCTG 840
 QY 841 GGCCTCACCGGGATAAGAACTGAAGTACTACATCCAGAGTCTTCTACACCGCGTCT 900
 DB 841 GGCCTCACCGGGATAAGAACTGAAGTACTACATCCAGAGTCTTCTACACCGCGTCT 900
 QY 901 ACCCACCGTATCCCGGGGTGGTGGATTCTCTACTCCGGCCCCCTTGGCTTGAGGCTGT 960
 DB 901 ACCCACCGTATCCCGGGGTGGTGGATTCTCTACTCCGGCCCCCTTGGCTTGAGGCTGT 960
 QY 961 ACAGTGGAGTACCGGGGTCATCTACCTTATGATGATGATGATGATGATGATGATGATG 1020
 DB 961 ACAGTGGAGTACCGGGGTCATCTACCTTATGATGATGATGATGATGATGATGATGATG 1020
 QY 1021 TTCAGAACTGGGCTTGTTCAGAGAGACAAAGGCTTCAGAGATTTGATTAATGAAG 1080
 DB 1021 TTCAGAACTGGGCTTGTTCAGAGAGACAAAGGCTTCAGAGATTTGATTAATGAAG 1080
 QY 1081 AGAAAAATAGAAAAATATTTTGTTCATATAGACCTAATGTTAGTACATAGCAGAAAC 1140
 DB 1081 AGAAAAATAGAAAAATATTTTGTTCATATAGACCTAATGTTAGTACATAGCAGAAAC 1140
 QY 1141 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1196
 DB 1141 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1196

RESULT 2
 AA167865
 ID AA167865 standard; cDNA; 1196 BP.
 XX
 AC AA167865;
 XX
 DT 13-MAR-2002 (first entry)
 XX
 DE Murine LIG46 polypeptide encoding cDNA.
 XX
 KW Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Stral3;
 KW anorectic; anabolic; antisense therapy; mouse; ss.
 XX
 OS Mus sp.
 XX
 EH Key Location/Qualifiers
 FT CDS 3..1194
 FT /tag= a
 FT /product= "LIG46"
 FT /note= "the coding sequence is specifically claimed"
 FT sig_peptide 3..98
 FT mat_peptide 99..1193
 FT /tag= b
 FT /tag= c
 XX US2001024808-A1.
 XX 27-SEP-2001.
 PD
 XX

PF 12-MAR-2001; 2001US-0804357.
XX
PR 29-OCT-1998; 98US-106378P.
PR 19-NOV-1998; 98US-0195896.
PR 10-SEP-1998; 98US-0150857.
XX
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI White D, Zhou J, Tartaglia LA;
XX
XX WPI; 2001-624963/72.
DR P-PSDB; AAG66115, AAG66116.
DR
XX An isolated polypeptide useful for treating a weight disorder and for
PT screening compounds that may inhibit protein activity comprises a
PT leptin-induced protein
XX
XX Claim 16; Fig 1: 46pp; Eng:lish.
XX
CC The invention relates to genes whose expression are induced by leptin.
CC Lig46 and Lig56 are novel leptin induced genes (lig), while four other
CC genes such as Tgfp, LRG-47, RCL0-11 and Stral3 have been previously
CC identified. The leptin induced proteins can be expressed by standard
CC recombinant methodology. The proteins and encoding polynucleotides may
CC be used in screening assays to identify compounds that may bind to it.
CC Administering a molecule, e.g., an antisense molecule, which reduces
CC expression of activity of protein selected from Lig46, Lig56, Tgfp,
CC LRP-47, RCL0-11, and Stral3 may be used to treat a weight disorder. The
CC method may further comprise administering leptin. The present sequence
CC represents a cDNA encoding a murine Lig46 protein.
XX
XX Sequence 1196 BP; 350 A; 269 C; 290 G; 287 T; 0 other;
XX

Query Match 100.0%; Score 1196; DB 23; Length 1196;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGATGAGTGTGGGCGTCGGAAGAGTCAAGTTGCTGGGCATCCTCGATGGCAAAATGTC 60
Db 1 AGATGAGTGTGGGCGTCGGAAGAGTCAAGTTGCTGGGCATCCTCGATGGCAAAATGTC 60
Qy 61 TCATTTATTTGATTTGGAAAGTCTCCAAACAGAGTACCCCAAGACAAATGGAAGGGAG 120
Db 61 TCATTTATTTGATTTGGAAAGTCTCCAAACAGAGTACCCCAAGACAAATGGAAGGGAG 120
Qy 121 GAGTAATATCCGGAAGAGAGTTCCTGGAAGCCACCCAGCCTCCCGGGCATATCTGGA 180
Db 121 GAGTAATATCCGGAAGAGAGTTCCTGGAAGCCACCCAGCCTCCCGGGCATATCTGGA 180
Qy 181 ACAGGGAACAGGAGAGCTGAACAGGTGTGTAATCCCATCTTTGAACAGGGTGGCCAATC 240
Db 181 ACAGGGAACAGGAGAGCTGAACAGGTGTGTAATCCCATCTTTGAACAGGGTGGCCAATC 240
Qy 241 AGACAGGGGAGCTAGGCACATCTCCAAACAGAGTCAACCTGAGCTATTGTGAACACAGACT 300
Db 241 AGACAGGGGAGCTAGGCACATCTCCAAACAGAGTCAACCTGAGCTATTGTGAACACAGACT 300
Qy 301 CGACGGTCATGACAGCTGTACAGAGATTTAATATCTGCGGACAGNTTAAAGACTTTC 360
Db 301 CGACGGTCATGACAGCTGTACAGAGATTTAATATCTGCGGACAGNTTAAAGACTTTC 360
Qy 361 TCTTGATTTGAGATCGCGGAATTAATCTGCTGCTTATAGATCAACCGAAGAAATGTGCAA 420
Db 361 TCTTGATTTGAGATCGCGGAATTAATCTGCTGCTTATAGATCAACCGAAGAAATGTGCAA 420
Qy 421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTGCCAGAGGCAAG 480
Db 421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTGCCAGAGGCAAG 480
Qy 481 CAATTCGGGAGTCTTGGGCGCGAGAAACCAACGCTAGGAGACACAGTAGTAGGGTCT 540
Db 481 CAATTCGGGAGTCTTGGGCGCGAGAAACCAACGCTAGGAGACACAGTAGTAGGGTCT 540

Qy 541 TCCTGIIGGCAAGACACACCCCCAGAGGACACACCACCTTGACCTTTTCGACATGCTTAAGT 600
Db 541 TCCTGIIGGCAAGACACACCCCCAGAGGACACACCACCTTGACCTTTTCGACATGCTTAAGT 600
Qy 601 TTGAGAGTGACAAGCACCAGGACATCCCTCATGTGGAACTATAGAGACACATCTTCAACC 660
Db 601 TTGAGAGTGACAAGCACCAGGACATCCCTCATGTGGAACTATAGAGACACATCTTCAACC 660
Qy 661 TGTCCCTGAAGGAAGTGTCTTTTCTTAGTGGGTGAGCATTCTCTGTCACACACAGAGT 720
Db 661 TGTCCCTGAAGGAAGTGTCTTTTCTTAGTGGGTGAGCATTCTCTGTCACACACAGAGT 720
Qy 721 TTGTCITCAAGGGCGATGATGACGTGTTTGTGAACACCCATCAGATCCTTAATTACTTGA 780
Db 721 TTGTCITCAAGGGCGATGATGACGTGTTTGTGAACACCCATCAGATCCTTAATTACTTGA 780
Qy 781 ATAGCTTATCCAAGAGCAAGCAAGGACTTGTTCATAGTGCAGTGATCCACCAATGCTG 840
Db 781 ATAGCTTATCCAAGAGCAAGCAAGGACTTGTTCATAGTGCAGTGATCCACCAATGCTG 840
Qy 841 GGCCTCACCGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACGGCGTCT 900
Db 841 GGCCTCACCGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACGGCGTCT 900
Qy 901 ACCCACCCTATGCCGGGGTGGTGATTCCTGTACTCGGCCCTTGCCTTGAGGCTGT 960
Db 901 ACCCACCCTATGCCGGGGTGGTGATTCCTGTACTCGGCCCTTGCCTTGAGGCTGT 960
Qy 961 ACAGTCCGACTAGCCGGGTCCATCTCTACCTATTGATGATGTTTATACGGGAATGTGCC 1020
Db 961 ACAGTCCGACTAGCCGGGTCCATCTCTACCTATTGATGATGTTTATACGGGAATGTGCC 1020
Qy 1021 TTCAGAAACTGGCCCTTCTTCCAGAGAGCACAAAGGCTTCAGGACATTTGATATTGAAG 1080
Db 1021 TTCAGAAACTGGCCCTTCTTCCAGAGAGCACAAAGGCTTCAGGACATTTGATATTGAAG 1080
Qy 1081 AGAAAAATAAGAAAAAATTTTGTTCCTATATAGACCTAATGTAGTACATAGCAGAAAAAC 1140
Db 1081 AGAAAAATAAGAAAAAATTTTGTTCCTATATAGACCTAATGTAGTACATAGCAGAAAAAC 1140
Qy 1141 CTCAGAGATGATTGATATCTGGTCTCAGTTCGCAAGTCCCTAAATTTAAATGCTGA 1196
Db 1141 CTCAGAGATGATTGATATCTGGTCTCAGTTCGCAAGTCCCTAAATTTAAATGCTGA 1196

RESULI 3
AAD46672
ID AAD46672 standard; cDNA; 1196 BP.
XX
XX AAD46672;
XX
XX 27-JAN-2003 (first entry)
XX
XX Murine Lig46 cDNA.
DE
XX
XX Lig46: leptin; low body weight; chromosomal mapping; tissue typing;
KW forensic biology; transgenic; gene therapy; antianorectic; murine;
KW gene; ss.
XX
XX Mus sp.
OS
XX Location/Qualifiers
FH 3..1196
FT /tag- a
FT /product- "Murine Lig46 protein"
FT sig_peptide 3..98
FT /tag- b
FT mat_peptide 99..1193
FT /tag- c
FT /product- "Murine mature Lig46 protein"
XX
XX W0200274905-A2.
XX

XX
DR WPI: 2000-161481/15.
DR P-PSDB: AAY79953.
XX
XX Mammalian EGGHEAD and BRAINIAc proteins which mediate cell to cell
PT adhesion and may be used to treat cancer, psoriasis and other skin
PT lesions and nervous system defects or diseases -
XX
PS Example 1: Page 24-25; 30pp: English.
XX
XX The present invention describes mammalian EGGHEAD and BRAINIAc proteins,
CC which mediate cell-to-cell adhesion. The mammalian proteins are similar
CC to the EGGHEAD and BRAINIAc genes of Drosophila melanogaster. BRAINIAc
CC and EGGHEAD genes regulate adhesion between epithelial cells and require
CC the presence of the Notch protein for this activity. Mutational
CC disruption of BRAINIAc, EGGHEAD or Notch results in the loss of follicle
CC epithelial cell adhesion, thereby reducing the efficiency of signalling
CC through other epithelial cell receptors such as the epidermal growth
CC factor receptor. The BRAINIAc and EGGHEAD proteins have cytosolic,
CC dermatological and neuroactive properties. As the BRAINIAc and EGGHEAD
CC proteins are involved in adhesion between epithelial cells, and as this
CC adhesion is required for cell viability, cell growth regulation and cell
CC fate specification, it is envisioned that wild type or mutant forms of cell
CC mammalian BRAINIAc and/or EGGHEAD can be used to alter epithelial cell
CC adhesion. This should be useful in treating many diseases which present
CC problems of cell viability, cell growth regulation and cell fate
CC specification. For example, these proteins, or active fragments or
CC analogues of these proteins and these genes can be used to treat
CC diseases such as cancer, psoriasis and other skin lesions, and nervous
CC system defects or diseases. The present sequence encodes the murine
CC BRAINIAc protein.
XX
SQ Sequence 2676 BP; 732 A; 562 C; 637 G; 744 T; 1 other;

Query Match 100.0%; Score 1196; DB 21: Length 2676;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGGCTCGAAGAGTCAAGTTCGCTGGGCATCCTGATGATGCAATGCT 60
DB 208 AGATGAGTGTGGGGCTCGAAGAGTCAAGTTCGCTGGGCATCCTGATGATGCAATGCT 267

QY 61 TCATTATTGTTGGAGTCTCCAAAACAGTAGTCCAAAGACAAAATGGAAAGGGAG 120
DB 268 TCATTATTGTTGGAGTCTCCAAAACAGTAGTCCAAAGACAAAATGGAAAGGGAG 327

QY 121 GAGTAAATATCCGAAACAGAGTTCGGAAGCCACCCAGCACTCCCGGGCATACTGGA 180
DB 328 GAGTAAATATCCGAAACAGAGTTCGGAAGCCACCCAGCACTCCCGGGCATACTGGA 387

QY 181 ACAGGGAACAGAGAGTGTACAGGTGGTACATCCCATCTTGAACAGGGTGGCCAAATC 240
DB 388 ACAGGGAACAGAGAGTGTACAGGTGGTACATCCCATCTTGAACAGGGTGGCCAAATC 447

QY 241 ACAGAGGGGAGCTAGCCACATCTCCAAACACAAGTCACTGAGCTATTGTGAACCACT 300
DB 448 ACAGAGGGGAGCTAGCCACATCTCCAAACACAAGTCACTGAGCTATTGTGAACCACT 507

QY 301 CCACGGTCTACAGCTGTGACAGATTTTAATAATCTCCCGGACAGATTTAAAGACTTTC 360
DB 508 CGACGGTCTACAGCTGTGACAGATTTTAATAATCTCCCGGACAGATTTAAAGACTTTC 567

QY 361 TCTTGTATTGAGTCCCGGAATTAATCTCGCTGCTTATAGATCAACCCGAGAAATGTGCAA 420
DB 568 TCTTGTATTGAGTCCCGGAATTAATCTCGCTGCTTATAGATCAACCCGAGAAATGTGCAA 627

QY 421 AGAAGCCCTTCTACTATTGTCGATTAAGTCCCTCATCTCCACATTTTCCGAGAGGCAAG 480
DB 628 AGAAGCCCTTCTACTATTGTCGATTAAGTCCCTCATCTCCACATTTTCCGAGAGGCAAG 687

QY 481 CAATTCCGGGAGTCTTGGGGCCGAGAAACCAACGTTAGGGAACGACAGTACTGAGGTCT 540
DB 688 CAATTCCGGGAGTCTTGGGGCCGAGAAACCAACGTTAGGGAACGACAGTACTGAGGTCT 747

QY 541 TCCTGTGGCAAGACACCCCGAGAGCAACACCCCTGACCTTTCGACATGCTTAAGT 600
DB 748 TCCTGTGGCAAGACACCCCGAGAGCAACACCCCTGACCTTTCGACATGCTTAAGT 807

QY 601 TTGAGAGTGAACAGACACCCAGAGCATCTCTCATGTGGAACACTATAGAGACACATTTCTCAACC 660
DB 808 TTGAGAGTGAACAGACACCCAGAGCATCTCTCATGTGGAACACTATAGAGACACATTTCTCAACC 867

QY 661 TGTCCCTCAAGGAAGTGTCTTTCTTAGGTGGGTGAGCACTTCTCTCCAGACGACAGT 720
DB 868 TGTCCCTCAAGGAAGTGTCTTTCTTAGGTGGGTGAGCACTTCTCTCCAGACGACAGT 927

QY 721 TTGCTTCAAGGGGATGATGACGTGTTTGTGAACCCATCATACATCCTTAATTAATCTGA 780
DB 928 TTGCTTCAAGGGGATGATGACGTGTTTGTGAACCCATCATACATCCTTAATTAATCTGA 987

QY 781 ATAGCTTATCCAAAGAGCAAGCAAGCAAGTGTTCATAGGTGAGTGTATCCACAATGCTG 840
DB 988 ATAGCTTATCCAAAGAGCAAGCAAGCAAGTGTTCATAGGTGAGTGTATCCACAATGCTG 1047

QY 841 GGCCTCACCGGATTAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACCGGCTCT 900
DB 1048 GGCCTCACCGGATTAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACCGGCTCT 1107

QY 901 ACCCACCGTATCCCGGGGTGGTGGATTCTCTACTCCGGCCCTTGCCTTGAGGCTGT 960
DB 1108 ACCCACCGTATCCCGGGGTGGTGGATTCTCTACTCCGGCCCTTGCCTTGAGGCTGT 1167

QY 961 ACAGTGGGAGTACCGGGTCCATCTCTACCTATTGATGATGTTTATACGGGAATGTGCC 1020
DB 1168 ACAGTGGGAGTACCGGGTCCATCTCTACCTATTGATGATGTTTATACGGGAATGTGCC 1227

QY 1021 TTCAGAACTGGCCCTTCTCCAGAGAGCAAGGCTTCAGGACATTTGATATTGAAG 1080
DB 1228 TTCAGAACTGGCCCTTCTCCAGAGAGCAAGGCTTCAGGACATTTGATATTGAAG 1287

QY 1081 AGAAAAATAAGAAAAATATTTTCTCTATATAGACCTAATGTTAGTACATAGCAGAAAAAC 1140
DB 1288 AGAAAAATAAGAAAAATATTTTCTCTATATAGACCTAATGTTAGTACATAGCAGAAAAAC 1347

QY 1141 CTCAGAGATGATGATATCTGCTCAGTTCAGTTCGAAAGTCTTAATTTAAATGCTGA 1196
DB 1348 CTCAGAGATGATGATATCTGCTCAGTTCGAAAGTCTTAATTTAAATGCTGA 1403

RESULT 5
AAZ87185
ID AAZ87185 standard; cDNA; 2676 BP.
XX
AC AAZ87185;
XX
DT 08-MAY-2000 (first entry)
XX
DE Murine Brainiac cDNA.
XX
XX Brainiac; murine; mammalian; expressed sequence tag; EST;
KW Drosophila melanogaster; Egghead; Notch; epithelial cell adhesion;
KW Fringe family; cell viability; growth regulation; cell fate; cancer;
KW psoriasis; skin lesion; nervous system disorder; developmental syndrome;
KW transgenic animal; drug screening; ss.
XX
OS Mus sp.
XX
XX Key Location/Qualifiers
FH 210..1403
ET CDS
ET /*tag- a
FT /product- "Murine Brainiac protein"
XX
PN CA2255109-A1.
XX
XX
PD 17-JUN-1999.
XX

PF 17-DEC-1998; 98CA-2255109.
 XX 17-DEC-1997; 97CA-2225126.
 XX (HSCR-) HSC RES & DEV LP.
 XX Egan; SE;
 XX WPI: 2000-148082/14.
 DR P-PSDB; AAY69697.
 XX
 PT New nucleic acids encoding a murine and human Brainiac protein. The cDNA
 PT for detecting somatic or germline DNA lesions which are responsible for
 PT developmental syndromes or diseases including cancer -
 XX
 PS Claim 5; Page 26-27; 40pp; English.
 XX
 CC This sequence represents cDNA encoding murine Brainiac protein. The cDNA
 CC was isolated from a mouse mammary gland cDNA library via the use of a
 CC probe generated via PCR from a variety of murine tissues. The PCR
 CC primers used for probe generation (AAZB7187-287188) were based on EST
 CC (expressed sequence tag) sequences with significant similarity to
 CC Drosophila melanogaster Brainiac. Drosophila Brainiac and Egghead
 CC proteins regulate adhesion between epithelial cells, this activity being
 CC dependent on the presence of Notch. Drosophila Brainiac is a secreted
 CC protein, and has sequence similarities with the Drosophila Fringe
 CC proteins. Brainiac/Egghead-mediated epithelial cell adhesion is required
 CC for cell viability, cell growth regulation and cell fate specification.
 CC Wild-type or mutant forms of mammalian Brainiac proteins may therefore be
 CC used to alter epithelial cell adhesion in a mammal. Mammalian Brainiac
 CC proteins, active fragments analogues, and nucleic acids may be used to
 CC treat diseases such as cancer, psoriasis and other skin lesions, and
 CC nervous system disorders. Mammalian Brainiac nucleic acids may also be
 CC used to detect somatic or germline DNA lesions which are responsible for
 CC developmental syndromes or diseases including cancer. The mammalian
 CC Brainiac proteins and fragments or its analogues are useful as antigens
 CC in immunoassays including enzyme-linked immunosorbent assays (ELISA),
 CC radioimmunoassays (RIA) and other non-enzyme linked antibody binding
 CC assays. Non-human transgenic animals comprising nucleotide sequences
 CC encoding human Brainiac protein (AAY69698) can be used as animal models
 CC for the study of mammalian Brainiac gene function, for the screening of
 CC candidate compounds and for the evaluation of potential therapeutic
 CC interventions.
 XX
 SQ Sequence 2676 BP: 732 A; 562 C; 637 G; 744 T; 1 other;
 Query Match 100.0%; Score 1196; DB 21; Length 2676;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 TCITGTATTGAGATGCCGAATTACTCGCTGCTTATAGATCAACGGAAGAAATGTCAA 420
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 568 TCITGTATTGAGATGCCGAATTACTCGCTGCTTATAGATCAACGGAAGAAATGTCAA 627
 QY 421 AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCATTCACATTTTGCACGAAGCGAAG 480
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 628 AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCATTCACATTTTGCACGAAGCGAAG 687
 QY 481 CAATTCGGGAGTCTTGGGCGCGAGAAACCAACCTAGGGAACCCAGACAGTAGTGAGGGTCT 540
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 688 CAATTCGGGAGTCTTGGGCGCGAGAAACCAACCTAGGGAACCCAGACAGTAGTGAGGGTCT 747
 QY 541 TCCTGTTGGGAGACACACCCCCAGAGACAAACCCCTGACCTTTCGGACATGCTTAAAGT 600
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 748 TCCTGTTGGGAGACACACCCCCAGAGACAAACCCCTGACCTTTCGGACATGCTTAAAGT 807
 QY 601 TTGAGGTGACAAGCACACGAGACATCCTCATGTGGAACTATAGAGACACATTCCTCAACC 660
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 808 TTGAGGTGACAAGCACACGAGACATCCTCATGTGGAACTATAGAGACACATTCCTCAACC 867
 QY 661 TGTCCCTGAAGGAGTCTCTTCTTCTTGTGGGTGAGCACATTCCTGTCCAGACGAGAGT 720
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 868 TGTCCCTGAAGGAGTCTCTTCTTCTTGTGGGTGAGCACATTCCTGTCCAGACGAGAGT 927
 QY 721 TTGTCTTCAGGGCGATGATGAGTGTCTTGTGAACCCCATCACATCCTTAATCTTGA 780
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 928 TTGTCTTCAGGGCGATGATGAGTGTCTTGTGAACCCCATCACATCCTTAATCTTGA 987
 QY 781 ATAGCTTATCCAAAGAGCAAGCAAGCAAGCTTGTTCATAGTGACGCTGATCCCAATGTGT 840
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 988 ATAGCTTATCCAAAGAGCAAGCAAGCAAGCTTGTTCATAGTGACGCTGATCCCAATGTGT 1047
 QY 841 GGCCTCACCGGATTAAGAACTGAAGTACTACATCCCAAGGCTTCTTACACCGGCTCT 900
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 1048 GGCCTCACCGGATTAAGAACTGAAGTACTACATCCCAAGGCTTCTTACACCGGCTCT 1107
 QY 901 ACCCAGCTATGCGGGGGTGGTGGATTCCTGTACTCCGGCCCTTGGCTTGAGGCTGT 960
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 1108 ACCCAGCTATGCGGGGGTGGTGGATTCCTGTACTCCGGCCCTTGGCTTGAGGCTGT 1167
 QY 961 ACAGTGGGACTAGCGGGTCCATCTTACCTTATGATGATGTTTATAGGGGAATGTGC 1020
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 1168 ACAGTGGGACTAGCGGGTCCATCTTACCTTATGATGATGTTTATAGGGGAATGTGC 1227
 QY 1021 TTCAGAAACTGGGCTTGTTCAGAGAAAGCACAAGAGCTTCAGACATTTGATATTGAAG 1080
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 1228 TTCAGAAACTGGGCTTGTTCAGAGAAAGCACAAGAGCTTCAGACATTTGATATTGAAG 1287
 QY 1081 AGAAAAATAGAAAAATATTGTTCTTATATAGACCTAATGTAGTACATAGCAGAAAC 1140
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 1288 AGAAAAATAGAAAAATATTGTTCTTATATAGACCTAATGTAGTACATAGCAGAAAC 1347
 QY 1141 CTCAGAGATGATTGATATCTGGTCTCAGTGTGCAAGCTCTTAATTTTAAATGCTGA 1196
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 1348 CTCAGAGATGATTGATATCTGGTCTCAGTGTGCAAGCTCTTAATTTTAAATGCTGA 1403
 RESULT 6
 AA167866
 ID AA167866 standard; cDNA; 1194 BP.
 XX
 AC AA-57866;
 XX
 DT 13-MAR-2002 (first entry)
 XX
 DE Murine IIG46 polypeptide coding sequence.
 XX
 KW Leptin; LIG46; LIG56; leptin induced gene; Tgfp; LRG-47; RC10-II; Stral3;
 KW anorectic; anabolic; antisense therapy; mouse; ss.
 XX
 OS Mus sp.
 XX

FH Key Location/Qualifiers
 FT CDS 1..1194
 FT /*tag= a
 FT /product= "LIG46"
 FT sig_peptide 1..96
 FT mat_peptide 97..1191
 FT /*tag= b
 FT /*tag= c
 XX US2001024808-A1.
 PN 27-SEP-2001.
 XX 12-MAR-2001; 2001US-0804357.
 XX 29-OCT-1998; 98US-106378P.
 PR 19-NOV-1998; 98US-0195896.
 PR 10-SEP-1998; 98US-0150857.
 XX (MILL-) MILLENNIUM PHARM INC.
 XX White D, Zhou J, Tartaglia LA;
 PI WPI; 2001-624963/72
 DR P-PSDB; AAG66115, AAG66116.
 XX An isolated polypeptide useful for treating a weight disorder and for
 PT screening compounds that may inhibit protein activity comprises a
 PT leptin-induced protein
 XX Claim 16; Fig 1; 46pp; English.
 XX The invention relates to genes whose expression are induced by leptin.
 CC LIG46 and LIG56 are novel leptin induced genes (LIG), while four other
 CC genes such as Tgfp, LRG-47, RC10-1f and Stral3 have been previously
 CC identified. The leptin induced proteins can be expressed by standard
 CC recombinant methodology. The proteins and encoding polynucleotides may
 CC be used in screening assays to identify compounds that may bind to it.
 CC Administering a molecule, e.g., an antisense molecule, which reduces
 CC expression of activity of protein selected from LIG46, LIG56, Tgfp,
 CC LRP-47, RC10-1f, and Stral3 may be used to treat a weight disorder. The
 CC method may further comprise administering leptin. The present sequence
 CC represents the murine LIG46 protein coding sequence.
 XX Sequence 1194 BP; 349 A; 269 C; 289 G; 287 T; 0 other;
 SQ Query Match 99.8%; Score 1194; DB 23; Length 1194;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 ATGAGTGTGGGGCTCGAAGAGTCAAGTTGCTGGGCATCCTGATGTCGAATGCTTTC 62
 DB 1 ATGAGTGTGGGGCTCGAAGAGTCAAGTTGCTGGGCATCCTGATGTCGAATGCTTTC 60
 QY 63 ATTTATTTGATTGTCGAAGTCTCCAAAACAGTAGCCAAAGACAAAATGGAAGGAGGA 122
 DB 61 ATTTATTTGATTGTCGAAGTCTCCAAAACAGTAGCCAAAGACAAAATGGAAGGAGGA 120
 QY 123 GTAATAATCCGAAGAGAGTCTTGGAAAGCCACCCAGCAGCTCCCGGGGCATCTGGAAAC 182
 DB 121 GTAATAATCCGAAGAGAGTCTTGGAAAGCCACCCAGCAGCTCCCGGGGCATCTGGAAAC 180
 QY 183 AGGACACAGAGAGTGTACAGAGTGTGATCCCATCTTGAACAGGTGGCCCAATCAG 242
 DB 181 AGGAAACAGAGAGTGTACAGAGTGTGATCCCATCTTGAACAGGTGGCCCAATCAG 240
 QY 243 ACAGGGGAGCTAGCCATCTCCAAACAGAGTCACTGAGCTATTGTGAACACGACTCG 302
 DB 241 ACAGGGGAGCTAGCCATCTCCAAACAGAGTCACTGAGCTATTGTGAACACGACTCG 300
 QY 303 ACGGTGATGACAGCTGTGACAGATTTTAATAATCTGCCGGGACAGATTTAAAGACTTTCTC 362
 DB 301 ACGGTGATGACAGCTGTGACAGATTTTAATAATCTGCCGGGACAGATTTAAAGACTTTCTC 360
 QY 363 TIGTATTGAGATGCCGGAATTAATCGCTGCTTATAGATCAACCCGAAGAAATGIGCAAG 422
 DB 361 TTGTATTGAGATGCCGGAATTAATCGCTGCTTATAGATCAACCCGAAGAAATGIGCAAG 420
 QY 423 AAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAGCA 482
 DB 421 AAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAGCA 480
 QY 483 ATTCCGGAGTCTTGGGGCCGAGAAACCAAGTAGGAAACCCAGACAGTGTAGGGTCTTC 542
 DB 481 ATTCCGGAGTCTTGGGGCCGAGAAACCAAGTAGGAAACCCAGACAGTGTAGGGTCTTC 540
 QY 543 CTGTTGGGCAAGACACCCCCAGAGGACCAACCCCTGACCTTTTCGGACATGCTTAAAGTTT 602
 DB 541 CTGTTGGGCAAGACACCCCCAGAGGACCAACCCCTGACCTTTTCGGACATGCTTAAAGTTT 600
 QY 603 GAGAGTCAAGACACCCAGGACATCCTCATGTGGAATATAGACACATCTTCAACCTG 662
 DB 601 GAGAGTCAAGACACCCAGGACATCCTCATGTGGAATATAGACACATCTTCAACCTG 660
 QY 663 TCCTGAAGGAAGTGTCTTTCTTAGTGGTGGTGGACATCTCTGTCCAGACGAGAGTTT 722
 DB 661 TCCTGAAGGAAGTGTCTTTCTTAGTGGTGGTGGACATCTCTGTCCAGACGAGAGTTT 720
 QY 723 GTCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCAGATCCTTAACTTGAAT 782
 DB 721 GTCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCAGATCCTTAACTTGAAT 780
 QY 783 AGCTTATCCAAAGACCAAGACATCTTCTATAGTGGTGGTGGACATCTTCAACATGCTGG 842
 DB 781 AGCTTATCCAAAGACCAAGACATCTTCTATAGTGGTGGTGGACATCTTCAACATGCTGG 840
 QY 843 CCTCACCGGATAGAAACTGAAGTACTACATCCCAAGTCTTCTACACGGCGTCTAC 902
 DB 841 CCTCACCGGATAGAAACTGAAGTACTACATCCCAAGTCTTCTACACGGCGTCTAC 900
 QY 903 CCACCGTATGCCGGGGTGGTGGATTCTGTACTCCGGCCCCCTTGCCTTGAGGCTGTAC 962
 DB 901 CCACCGTATGCCGGGGTGGTGGATTCTGTACTCCGGCCCCCTTGCCTTGAGGCTGTAC 960
 QY 963 AGTCGACTAGCCGGGTCCATCTCTACCTATTGATGATGCTTTATACGGGAATGTGCCT 1022
 DB 961 AGTCGACTAGCCGGGTCCATCTCTACCTATTGATGATGCTTTATACGGGAATGTGCCT 1020
 QY 1023 CAGAACTGGGCCCTTGTTCAGAGAACCAAGGCTTCAGGACATTTGATATTGAAGAG 1082
 DB 1021 CAGAACTGGGCCCTTGTTCAGAGAACCAAGGCTTCAGGACATTTGATATTGAAGAG 1080
 QY 1083 AAAAATAGAAATAATTTGTTCTATAGACCTAATGTTAGTACATACGAAACCT 1142
 DB 1081 AAAAATAGAAATAATTTGTTCTATAGACCTAATGTTAGTACATACGAAACCT 1140
 QY 1143 CAAGAGATGATTGATATCTGCTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 1196
 DB 1141 CAAGAGATGATTGATATCTGCTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 1194
 RESULT 7
 ABV75081
 ID ABV75081 standard; DNA; 1260 BP.
 XX ABV75081;
 XX 19-FEB-2003 (first entry)
 XX Murine beta3nT gene polynucleotide sequence.
 XX Beta3nT; transgenic; beta-1,3-N-acetylglucosaminyltransferase;
 KW disease system; gene; mouse; ds.
 XX Mus sp.
 OS

XX Key Location/Qualifiers
 FT CDS 30..1223
 FT /*tag= a
 FT /product= "beta3GnT"
 PN WC202279413-A2.
 XX 10-OCT-2002.
 XX 29-MAR-2002; 2002WO-US09645.
 XX 29-MAR-2001; 2001US-280706P.
 PR 28-MAR-2002; 2002US-0280706.
 XX (DELTA-) DELTAGEN INC.
 XX Leviten MW, Phillips R;
 XX WPI; 2003-067437/06.
 DR P-PSDB; AB882648.
 XX New transgenic mouse comprising a disruption in a beta3GnT gene, as in
 PT vivo model to study various disease states or conditions in which
 PT beta3GnT may be implicated or involved, such as abnormal cell growth,
 PT cancer and metastasis -
 XX Disclosure; Fig 1; 55pp; English.
 XX The invention relates to a transgenic mouse comprising a disruption in a
 CC beta3GnT (beta-1,3-N-acetylglucosaminyltransferase) gene, where there is
 CC no native expression of beta3GnT gene. The transgenic mice may be used
 CC as in vivo model to study various disease states or conditions in which
 CC beta3GnT may be implicated or involved, such as abnormal cell growth,
 CC cancer and metastasis, and to evaluate various treatments or to identify
 CC agents for treating disease states or conditions, such as anxiety or
 CC depression. Animal-based disease systems may be used to identify
 CC compounds capable of ameliorating disease symptoms, as test substrates
 CC for the identification of drugs, pharmaceuticals, therapies and
 CC interventions that may be effective in treating a disease or other
 CC phenotypic characteristic of the animal. The present sequence represents
 CC a mouse beta3GnT polynucleotide sequence.
 XX
 SQ Sequence 1260 BP; 370 A; 282 C; 307 G; 301 T; 0 other;
 Query Match 99.7%; Score 1192.8; DB 25; Length 1260;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1194; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AGATGAGTGTGGGCTCGAAGAGTCAAGTTCCTGGGCATCCTGATGATGGCAATGTCT 60
 DB 28 AGATGAGTGTGGGCTCGAAGAGTCAAGTTCCTGGGCATCCTGATGATGGCAATGTCT 87
 QY 61 TCATTATTATTGTTGGAGTCTCCAAAACAGTAGCCACACAAAATGGAAAGGAG 120
 DB 88 TCATTATTATTGTTGGAGTCTCCAAAACAGTAGCCACACAAAATGGAAAGGAG 147
 QY 121 GAGTAATAATCCGAAAGAGAGTCTGGAAGCCACCCAGCTCCCGGGGCATCTGGA 180
 DB 148 GAGTAATAATCCGAAAGAGAGTCTGGAAGCCACCCAGCTCCCGGGGCATCTGGA 207
 QY 181 ACAGGGAACAGGAGAGTCTGACAGTGTGACATCCCATCTGACAGGGTGGGCAATC 240
 DB 208 ACAGGGAACAGGAGAGTCTGACAGTGTGACATCCCATCTGACAGGGTGGGCAATC 267
 QY 241 AGACAGGGAGCTAGCCATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACT 300
 DB 268 AGACAGGGAGCTAGCCATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACT 327
 QY 301 CGACGGTCATGACAGCTGTGACAGATTTTAATAATCTGCCGACAGATTTAAGACTTTC 360
 DB 328 CGACGGTCATGACAGCTGTGACAGATTTTAATAATCTGCCGACAGATTTAAGACTTTC 387

QY 361 TCTGTATTGAGATGCCGGAATTAATCTCGCTGCTTATAGATCAACCCGAAGAAATGICAA 420
 DB 388 TCTGTATTGAGATGCCGGAATTAATCTCGCTGCTTATAGATCAACCCGAAGAAATGICAA 447
 QY 421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTGCCAGAGGCAAG 480
 DB 448 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTGCCAGAGGCAAG 507
 QY 481 CAATTGGGAGTCTTTGGGGCCGAGAAACCAAGCTAGGGAACGACAGACTAGTGGGGTCT 540
 DB 508 CAATTGGGAGTCTTTGGGGCCGAGAAACCAAGCTAGGGAACGACAGACTAGTGGGGTCT 567
 QY 541 TCTGTGTGGCAAGACACCCCCAGAGGACACACCCCTGACCTTTTCGGACAIGTCTAAGT 600
 DB 568 TCTGTGTGGCAAGACACCCCCAGAGGACACACCCCTGACCTTTTCGGACAIGTCTAAGT 627
 QY 601 TTGAGAGTGCACAGCAGCAGGACATCCTCATGTGTGAACATATAGAGACACATCTTCAACC 660
 DB 628 TTGAGAGTGCACAGCAGCAGGACATCCTCATGTGTGAACATATAGAGACACATCTTCAACC 687
 QY 661 TGTCCCTGAAGGAAGTGTCTTTCTTAGTGGGTGAGCACTTCTCTGTCACAGCAGCAGT 720
 DB 688 TGTCCCTGAAGGAAGTGTCTTTCTTAGTGGGTGAGCACTTCTCTGTCACAGCAGCAGT 747
 QY 721 TTGCTTTCAGGGCGGATGATGACGTGTTGTGAACACCCATCACATCTTAACTACTGA 780
 DB 748 TTGCTTTCAGGGCGGATGATGACGTGTTGTGAACACCCATCACATCTTAACTACTGA 807
 QY 781 ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGGTGACGTGATCCACAATGCTG 840
 DB 808 ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGGTGACGTGATCCACAATGCTG 867
 QY 841 GGCTCACCAGGATAGAAACTGAACTACTACATCCAGAGTCTTCTACACCCGGCTCT 900
 DB 868 GGCTCACCAGGATAGAAACTGAACTACTACATCCAGAGTCTTCTACACCCGGCTCT 927
 QY 901 ACCACCGTATGCGGGGTGGTGGATTCCTGATCGGGCCCCCTTGCCTTGAGGGTGT 960
 DB 928 ACCACCGTATGCGGGGTGGTGGATTCCTGATCGGGCCCCCTTGCCTTGAGGGTGT 987
 QY 961 ACAGTGGCACTACCGGGTCCATCTCTACCTTATGATGATGTTTATACGGGAATGTGCC 1020
 DB 988 ACAGTGGCACTACCGGGTCCATCTCTACCTTATGATGATGTTTATACGGGAATGTGCC 1047
 QY 1021 TTCAGAACTGGGCCCTTGTTCAGAGAGACACAAAAGGCTTCAGGACATTTGATATTGAAG 1080
 DB 1048 TTCAGAACTGGGCCCTTGTTCAGAGAGACACAAAAGGCTTCAGGACATTTGATATTGAAG 1107
 QY 1081 AGAAAAATAGAAAATATTTGTTCCATATAGACCTAAITGTTAGTACATAGCAGAAAAC 1140
 DB 1108 AGAAAAATAGAAAATATTTGTTCCATATAGACCTAAITGTTAGTACATAGCAGAAAAC 1167
 QY 1141 CTCAGAGATGATTGATATCTGTGCTCAGTTGCAAAAGTCCCTAAATTTAAATGCTGA 1196
 DB 1168 CTCAGAGATGATTGATATCTGTGCTCAGTTGCAAAAGTCCCTAAATTTAAATGCTGA 1223
 RESULT 8
 AAA-2716
 ID AAA12716 standard; cDNA; 1707 BP.
 XX AC AAA12716;
 XX AC AAA12716;
 DT 25-JUL-2000 (first entry)
 XX cDNA encoding human LIG46 polypeptide.
 DE Leptin; LIG46; body weight; leptin inducible gene; obesity; cachexia;
 KW LIG56; Tgtp; LRG-47; RC10-II; Stral3; ss.
 XX Homo sapiens.
 XX OS
 XX FH Key Location/Qualifiers

```

FI CDS      245..1439
FT /*tag= a
FX /product= "LIG46"
XX
XX WO200015826-A2.
XX
XX 23-MAR-2000.
XX
XX 10-SEP-1999; 99WO-US20722.
XX
XX 10-SEP-1998; 98US-0150857.
XX 29-OCT-1998; 98US-0106378.
XX 19-NOV-1998; 98US-0195896.
XX 15-APR-1999; 99US-0292228.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX White D. Zhou J. Tartaglia LA;
XX
XX WPI; 2000-271461/23.
XX P-PSDB; AAY84683.
XX
XX Method for determining compounds which modulate body weight and can be
XX used to treat e.g. obesity comprises measuring the activity of leptin;
XX Inducible genes
XX
XX Disclosure; Fig 7; 123pp; English.
XX
XX The present sequence encodes a leptin induced LIG46 polypeptide. The
XX specification describes a method for determining whether a compound can
XX be used to modulate body weight by measuring the activity of leptin
XX inducible genes, such as LIG46. The method can be used to specifically
XX detect LIG46 and LIG56 nucleic acid molecules relative to other nucleic
XX acid molecules encoding galactosyltransferases or GTP-binding proteins.
XX They can also be used in diagnostic assays to identify the presence or
XX absence of a genetic lesion or mutation characterized by aberrant
XX modification or mis-regulation of the genes or aberrant post-
XX translational modification of the proteins. LIG46 and LIG56 proteins
XX and nucleic acid molecules can be used to treat obesity or cachexia.
XX LIG46 and LIG56 antagonists are used to treat obesity and LIG46 and
XX LIG56 agonists are used to treat low body weight. The leptin inducible
XX genes i.e. LIG46, LIG56, Tgtp, LRG-47, RC10-II and Stral3 can be used
XX to determine whether a compound modulates body weight and can then be
XX used to treat obesity or cachexia or low body weight.
XX
XX Sequence 1707 BP: 452 A; 403 C; 435 G; 400 T; 17 other;
XX
XX Query Match      75.8%; Score 906.4; DB 21; Length 1707;
XX Best Local Similarity 84.9%; Pred. No. 5.6e-270;
XX Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
XX
XX 1 AGATGAGTGTGGGCGCTCGAAGAGTCAGTTCCTGGGCGATCCCTGATGTCGAAATGTCT 60
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 244 AAATGAGTGTGGGCGCTCGAAGATTAAGTTGTTGGGTATCCTGATGTCGAAATGTCT 303
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 61 TCATTATTGTTGTAAGTCTCCAAAACAGTACGACAGCAAAAATGGAAGGAG 120
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 304 TCATTATTATTATGAGTCTCCAAAACAGTACGACAGCAAAAATGGAAGGAG 363
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 121 GAGTAAATATCCGAAAGAGAGTCTTGAAGCCAGCCAGCATCCCGGGCGATCTGGA 180
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 364 AAGTAAATATCCGAAAGAGAGTCTTGAAGAGATATCTACCCCTCCCGGAGCATCTGGA 423
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 181 ACAGGGAACAGGAGAGTCTGACAGGTGTACATCTTCAACAGGCTGCCAATC 240
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 424 ACCGAGAGCAGAGAGTCTGACCGGAGTACACCCCATCTCTGAGCATCTGACCAAC 483
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 241 AGACAGGGGAGTACCCATCTCCAAACAGTACCTGAGCTATTGTGAACAGACT 300
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 484 ACACGGGGGAGCGGCGAGGCTCTCCAAATATAGCCATCTGAACCTGCGAAGCTGACC 543
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 301 CGACGGTCTACAGCTGTGACAGATTTTAAATATCTGCGGAGACAGATTTAAAGACTTC 360
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

544 TGAGGTCACGTCGGTGGTTACGGGTTTAAACAACCTTGCAGGACAGATTTAAAGACTTTC 603
361 TCTTGTATTTAGATGCGGGAATTTACTCGCTGTATAGATCAACGCAAGAAATGTGCRA 420
604 TGCTGTATTTAGATGCGGCAATTTACTCGCTGTATAGATCAGCGGGAATGTGTGCAA 663
421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTTCCACATTTTGCAGAGCAAG 480
664 AGAAGCCCTTCTTGTGCTGGGATTAAGTCCCTCACTCCACATTTTGCAGAGCAAG 723
481 CAATTCGGGAGTCTTGGGGCGGAGAAACCAACAGTAGGAAACAGACAGTAGTCTGAGGTCT 540
724 CAATTCGGGAAATCTTGGGGCGGCAAGAAAGCAACGAGGAAACCAAGGAGTGTGCGAGTCT 783
541 TCCTGTTGGGCAAGACACACCCCGGAGGACCAACACCTGACCTTTCGGACATCTTAAAGT 600
784 TCCTGCTGGGCGAGACACCCCGGAGGACCAACACCCCGGACCTTTCAGATATGCTGAAT 843
601 TTGAGAGTGACAGACACAGGACATCTCATGTGGAACTATAGAGACACATTTTCAACC 660
844 TTGAGAGTGAGAGCAACCAAGACATCTTATGTGGAACCTACAGAGACATTTCTTCAACT 903
661 TCTCCCTCAAGGAAGTCTGTTTCTTAGTGGGTGAGCACTTCTGTCAGAGCAGAGAGT 720
904 TGTCTCTGAAGGAAGTCTGTTTCTCAGGTGGGTGAAGTACTTCTGCCAGACACATGAGT 963
721 TTGCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCCATCAATCCCTTAATTAAGTGA 780
964 TTGTTTCAAGGGCGATGACCATGTTTGTGAACACCCCATCAATCCCTGAATTAAGTGA 1023
781 ATAGCTTATCAAGAGCAACCCCAAGACATTTGTTATAGGTGAGCTGATCCCAATGCTGT 840
1024 ATAGTTTATCAAGACCAACCAAGATCTTTCATAGGTGATGATCCCAATGCTGT 1083
841 GGCCTCACCGGGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACGGGCTGT 900
1084 GACCTCATCGGATGAAGAGCTGAAGTACTACATCCCAAGAGTGTGTTACTCTGGCTCT 1143
901 ACCCACTATCGCGGGGTGGTGGATTCCTGTACTCGGGCCCTTGCCTTGAAGCTGT 960
1144 ACCCACTATCGCGGGGTGGTGGATTCCTGTACTCGGGCCCTTGCCTTGAAGCTGT 1203
961 ACAGTGCCATGAGCGGTGCTTCTACCTTATGATGATGTTTATACGGGATGTC 1020
1204 ACCATATCACTGACAGGTGCTTCTACCTTATGATGATGTTTATACGGGATGTC 1263
1021 TFCAGAACTGGGCTTGTTCAGAGAGCAACAAAGGCTTCAGACATTTGATTAAG 1080
1264 TTCAGAACTGGGCTTGTTCAGAGAGCAACAAAGGCTTCAGACATTTGATTAAG 1323
1081 AGAAAAATAAGAAAAATATTGTTCTATATAGACCTAATGTTAGTACATAGCAGAAAC 1140
1324 AGAAAAATAAGAAAAATATTGTTCTATATAGACCTAATGTTAGTACATAGTAAAC 1383
1141 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1196
1384 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1439

RESULT 9
AAI67869
ID AAI67869 standard; cdna; 1707 BP.
XX
XX AAI67869;
AC
XX
XX 13-MAR-2002 (first entry)
DT
XX
XX Human LIG46 polypeptide encoding cdna.
DE
XX
XX Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Stral3;
XX anorectic; anabolic; antisense therapy; human; ss.
XX
XX Homo sapiens.
OS

```

XX FH Key Location/Qualifiers
 XX FT CDS 246..1439
 XX FT /*tag= a
 XX FT /product= "LIG46"
 XX
 PN US2001024808-A1.
 XX
 XX 27-SEP-2001.
 XX
 XX 12-MAR-2001; 2001US-0804357.
 XX
 XX 29-OCT-1998; 98US-106378P.
 XX
 XX 19-NOV-1998; 98US-0195896.
 XX
 XX 10-SEP-1998; 98US-0150857.
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 XX
 XX White D, Zhou J, Tartaglia LA;
 XX
 XX WPI; 2001-624963/72.
 XX
 XX P-PSDB; AAG66118.
 XX
 XX An isolated polypeptide useful for treating a weight disorder and for
 XX screening compounds that may inhibit protein activity comprises a
 XX leptin-induced protein
 XX
 XX Example 2: Fig 7; 46pp; English.
 XX
 XX The invention relates to genes whose expression are induced by leptin.
 XX LIG46 and LIG56 are novel leptin induced genes (LIG), while four other
 XX genes such as Tgfp, LRG-47, RC10-II and Stral3 have been previously
 XX identified. The leptin induced proteins can be expressed by standard
 XX recombinant methodology. The proteins and encoding polynucleotides may
 XX be used in screening assays to identify compounds that may bind to it.
 XX Administering a molecule, e.g., an antisense molecule, which reduces
 XX expression of activity of protein selected from LIG46, LIG56, Tgfp,
 XX LRG-47, RC10-II, and Stral3 may be used to treat a weight disorder. The
 XX method may further comprise administering leptin. The present sequence
 XX represents a cDNA encoding a human LIG46 protein.
 XX
 XX Sequence 1707 BP: 452 A; 403 C; 435 G; 400 T; 17 other;
 XX
 XX Query Match 75.8%; Score 906.4; DB 23; Length 1707;
 XX Best Local Similarity 84.9%; Pred. No. 5.6e-270;
 XX Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
 QY 1 AGATGAGTGTGGGGTCGAGAGCTCAAGTTGCTGGGCATCTTGATGATGAGCAATGTCT 60
 DB 244 AATGAGTGTGGACGTGCGAAGATAAAGTTGTTGGGTATCTCTGATGATGGCAATGTCT 303
 QY 61 TCATTTATTTGATTTGGAAGTCTCCAAAACAGTAGCCAGACAAAATGSAAGGGAG 120
 DB 304 TCATTTATTTATTTGAGTCTCCAAAAGCAGTAGCCAGACAAAATGSAAGGGAG 363
 QY 121 GAGTAATATCCCAAGAGAGAGTTCTGGAGGCCACCCAGCAGCTCCCGGGCATCTGGA 180
 DB 364 AAGTAATATATCCCAAGAGAGAGTTCTGGAGAGATATCTACCTCCCGGAGGCACTGGA 423
 QY 181 ACAGGAAACAGAGAGAGTGAACAGTGGTACAATCCCATCTTGAACAGGTGGCCCAATC 240
 DB 424 ACCGAGACAGAGAGAGTGAACCGGAGTAGACACCCCATCTCTGAGCACTGACCAACC 483
 QY 241 AGACAGGGAGTAGCCACATCTCCAAACACAGTACCTGAGTATTTGTAACAGAGACT 300
 DB 484 AGACAGGGGAGGGCGGAGGCTCTCCAAATATAAGCCATCTGAACCTACTGCGAAGCTGACC 543
 QY 301 CGACGGTCTATGACAGCTGTGACAGATTTTAATATCTGCGGACAGATTTAAGACATTC 360
 DB 544 TGAGGGTCACGTCGGTGGTGTACGGGTTTAAACAACCTTCGCGACAGATTTAAGACATTC 603
 QY 361 TCTTCTATTTGAGATGCGGAGTACTCGCTGCTTATAGATCAACCGCAAGAAATGTGCAA 420

DB 504 TGCTGATTTGAGATGCCGCAATTTTACTGCTTATAGATCAGCGCGGATAGTGTGCAA 663
 QY 421 AGAAGCCCTTCTTACTATTGGGATAAAGTCCCTCATTCACATTTTGCAGAAGCAAG 480
 DB 664 AGAAACCTTCTTCTGCTGGCGATTAAAGTCCCTCACTCCACATTTTGCAGAAGCAAG 723
 QY 481 CAATTCGGGAGTCTTGGGGCCGAGAACCAACAGTAGGGAACAGACAGATAGTGGAGTCT 540
 DB 724 CAATCCGGGAATCTTGGGGCCAAAGAACCAACAGTAGGGAACCAACAGTGGTGGAGTCT 783
 QY 541 TCCTGTTGGCAAGACACCCAGAGAGAACCAACCACTTCCGACATCTTCCGACATCTTAAGT 600
 DB 784 TCCTGTTGGCCAGACACCCAGAGAGAACCAACCACTTCCGACATCTTCCGACATCTTAAGT 843
 QY 601 TTGAGAGTGACAGACACCCAGAGACATCTCATGTGGAATATAGAGACACATCTTCAACC 660
 DB 844 TTGAGAGTGAGAACCAACAGACATCTTATGTGGAATATAGAGACACATCTTCTCAACT 903
 QY 661 TGTCCTTGAAGAGTGTCTTCTTAGTGGGTGAGCAGTCTCTCTCCAGACGAGAGT 720
 DB 904 TGCTCTGAAGAGTGTCTTCTCAGGTGGGTAAAGTCTCTCTCCAGACAGACATGAGT 963
 QY 721 TTGCTTTCAAGGGCGATGATGACGTCTTGTGAACACCCATCACATCTTAAATCTTGA 780
 DB 964 ITGTTTTCAAGGGCGATGACGATGTTTTGTGAACACCCATCACATCTTGAATCTTGA 1023
 QY 781 ATAGCTTATCCAGAGACCAAGACCTTGTTCATAGTGGAGTGTATCCCAATGCTG 840
 DB 1024 ATAGTTTATCCAGACCAAGACCAAGATCTTTCATAGTGGATGATGCCAATGCTG 1083
 QY 841 GGCTCACCGGATTAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACCGGCTCT 900
 DB 1084 GACCTCATCGGATAAAGACTGAAGTACTACATCCCAAGAGTGTGTTACTCTGGCTCT 1143
 QY 901 ACCACCGTATCCCGGGGTGGTGGATTCTGTACTCCGCGCCCTTGGCTTCAGGCTGT 960
 DB 1144 ACCACCGTATCCAGGGGGAGGGGGTCTCTACTCCGCGCCCTTGGCTTCAGGCTGT 1203
 QY 961 ACAGTGGCACTAGCCGGGTGCTCATCTCTACCTCTATGATGATGTTTATACGGGAATGCGC 1020
 DB 1204 ACCATATCACTGACCAAGTCTCTACCCCATGATGACGTTTATCTGGAATGCGC 1263
 QY 1021 TTCAGAACTGGGCTTGTCCAGAGAGACAAAGGCTTCAGAGACATTTGATTTGAAG 1080
 DB 1264 TTCAGAACTGGGCTTGTCCAGAGAGACAAAGGCTTCAGAGACATTTGATATCGAG 1323
 QY 1081 AGAAAAATAAGAAAAATATTGTCCTATATAGACCTAAATGTTAGTACATAGCAGAAAAAC 1140
 DB 1324 AGAAAAACAAAAATAACATCTGCTCTTATGATCTGATGATGATGATGATGATGATGAT 1383
 QY 1141 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1196
 DB 1384 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1439

RESULT 10

AA046673
 ID AAD46673 standard: cDNA; 1707 BP.

XX AAD46673;

DT 27-JAN-2003 (first entry)

XX Human LIG46 cDNA.

XX LIG46; leptin; low body weight; chromosomal mapping; tissue typing;
 KW forensic biology; transgenic; gene therapy; antianorectic; human;
 KW chromosome 2; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH 246..1439
 FT CDS

```

FT      /*tag= a
FT      /product= "Human LIG46 protein"
FT      sig_peptide 246...341
FT      *tag= b
FT      mat_peptide 342...1436
FT      /*tag= c
FT      /product= "Human mature LIG46 protein"
XX      WC200274905-A2.
XX      26-SEP-2002.
XX      20-NOV-2001; 2001WO-US43345.
XX      21-NOV-2000; 2000US-0717778.
XX      (MILL-) MILLENNIUM PHARM INC.
XX      White DW, Zhou J, Tartaglia LA, Stricker-krongrad A, Clausen H;
XX      WPI; 2002-759886/82.
XX      P-PSDB: AAE29093.
XX      New isolated nucleic acid molecules, designated as leptin-induced genes
XX      46 (LIG46), useful for treating a subject having a disorder
XX      characterized by undesirable level of LIG46 expression or activity.
XX      such as low body weight
XX      Claim 15; Fig 5; 90pp; English.
XX      The invention relates to LIG46, a gene whose expression is induced by
XX      leptin. LIG46 DNA and protein are useful in treating a subject having
XX      a disorder characterized by undesirable level of LIG46 expression or
XX      activity, such as low body weight. They are also useful in a screening
XX      assay, chromosomal mapping, tissue typing and forensic biology. The
XX      probes based on the LIG46 nucleotide sequence are useful for detecting
XX      transcripts or genomic sequences encoding the same or related proteins.
XX      The LIG46 polypeptides are useful as immunogens for raising anti-LIG46
XX      antibodies. The host cells are useful for producing non-human transgenic
XX      animals. LIG46 DNA is used in gene therapy. The present sequence is
XX      human LIG46 cDNA. LIG46 gene is located at chromosome 2.
XX      Sequence 1707 BP; 452 A; 403 C; 435 G; 400 T; 17 other:
XX      Query Match 75.8%; Score 906.4; DB 24; Length 1707;
XX      Best Local Similarity 84.9%; Pred. No. 5.6e-276;
XX      Matches 1045; Conservative 0; Mismatches 181; Indels 9; Gaps 0:
OY      1 AGATGAGTGTGGGGCGTCAAGAGTCAAGTTGCTGGGCATCCTGATGATGCAAAATGTCT 60
Db      244 AATGAGTGTGGAGCTCGAAGATAAAGTTGTTGGGTATCCTGATGATGCAAAATGTCT 303
OY      61 ICATTTATTATTGTTGGAGTCTCCAAAACAGTAGCCAAAGAAAATGGAAGGGAG 120
Db      304 TCATTTATTATTATGGAAGTCTCCAAAGCAGTAGCCAAAGAAAATGGAAGGGG 363
OY      121 GASTAATAATCCGAAAGAGAGTTCTTGAAGCCACCCAGCAGTCCCGGGCATCTATGGA 180
Db      364 AAGTAATAATACCAAGAGAGAGTTCTTGAAGATACTAGCCCTCCCGAGGCACTATGGA 423
OY      181 ACAGGGAACAGAGAGCTGACAGGTGGTACAAATCCCATCTTGAACAGGTGGCCAAATC 240
Db      424 ACCGAGAGCAAGAGAGCTGAAACCCGAGTAGCAACCCCATCTGAGCATGTGACCAACC 483
OY      241 AGACAGGGGAGCTAGCCATCTCCAAACACAGTCACTGAGCTATTGTGAACACAGACT 300
Db      484 AGACGGGGAGGGCGGCGCAGGCTCTCCAAATATAGGCATCTGACTCTGCAACCTGACC 543
OY      301 CGAGGCTCATGACAGCTGTGACAGATTTTAAATATCTCCCGACAGATTTAAAGACTTTC 360
Db      544 TGAGGGTCACTGCTGGTGGTACGGGTTTAAACAACTTCCCGACAGATTTAAAGACTTTC 603
OY      361 TCTTGTATTGTAGATCCCGAATTACTCGCTCTTAAGATCAACCCGAGAAATGTGCAA 420

```

```

Db      604 TGCTGATTTGAGATGCCGCAATTATTTCACATGCTATATAGATCAGCCGGATAGTGTCAA 663
OY      421 AGAAGCCCTTCTACTATTGGGATAAAGTCCCTCATTTCCACATTTTGGCCAGAGGAAG 480
Db      664 AGAAACCTTCTTGTGTTGGCGATTAAAGTCCCTCACTCCACATTTTGGCAGAGGAAG 723
OY      481 CAATTCCGGGAGTCTTGGGGCGAGAAACCAACGTAAGGGAACACACAGTAGTAGGGTCT 540
Db      724 CAATCCGGGAATCTTGGGGCCAGAAAGACACGAGGGAACCAACGGTGGTGGAGTCT 783
OY      541 TCTGTGGGCAAGACACCCCGAGAGAAACACCCCTGACCTTTCGGACATGCTTAAGT 600
Db      784 TCCGTGTCGGCAGACACCCCGAGAGAAACCAACCCCGAGCTTTCAGATATGCTGAAT 843
OY      601 TTGAGAGTGACAAAGCAGGAGACATCTCATGTGGAACATATAGAGACACATCTTCAACC 660
Db      844 TTGAGAGTGAGAAAGCAGCAAGACATCTTATGTGGAACATACAGAGACATTTCTTCACT 903
OY      661 TGTCCCTGAAGGAAGTGTCTTGTAGTGGGTGAGCACTTCTCTGAGTGCAGTATCCACAATGCTG 720
Db      904 TGTCTCTGAAGGAAGTGTCTTCTCAGGTGGTAACTTCTCTGAGTGCAGTATCCACAATGCTG 963
OY      721 TTGTCTTCAAGGGCGATGATGACGTGTTGTGAACACCCATCAGATCTCTTAATTAAGTGA 780
Db      964 TTGTTTCAAGGGCGATGACGATGTTTGTGAACACCCATCAGATCTCTGAATTAAGTGA 1023
OY      781 ATAGCTTATCCAAAGAGCAAGCAAGACTTGTTCATAGTGCAGTGCATCCACAATGCTG 840
Db      1024 ATAGTTTATCCAAAGACCAAGCAAGATCTCTTCTATAGTGTGATGATCCACAATGCTG 1083
OY      841 GGCCTCACCGGATGAAGAACTGAAGTACTACTATCCAGAGAGTCTTCTACACCGGCTCT 900
Db      1084 GACCTCATCGGATGAAGAGCTGAAGTACTATATCCAGAGAGTGTGTTACTCTGGCCCTCT 1143
OY      901 ACCACCGTATCCCGGGGTGGTGGATTCCTGTACTCCGGCCCTTCCCTTGAGGGTGT 960
Db      1144 ACCACCCCTATCAGGGGAGGGGGTCTCTACTCCGGCCCTGAGGGTGT 1203
OY      961 ACAGTGCAGTACGCGGGTCCATCTCTACCTTATGATGATGTTTATACGGGAATGTGCC 1020
Db      1204 ACCATATCACTGACCAAGTCCATCTCTACCCCAITGTAGTACGTTATATGATGTGCC 1263
OY      1021 TTCAGAACTGGGCTTGTTCAGAGAGACAAAGGCTTCAGGACATTTGATATGAAG 1080
Db      1264 TTCAGAACTGGGCTTGTTCAGAGAGACAAAGGCTTCAGGACATTTGATATGAGG 1323
OY      1081 AGAAATAAGAAAAATATTTCTTATATAGACCTAATGTTAGTACATAGCAAAAC 1140
Db      1324 AGAAATAAGAAAAATATCTCTCTCTATGATGATGTTAGTACATAGTAAAC 1383
OY      1141 CTCAGAGATGATGATATCTGCTCAGTTCGAAAGTCTTAATTTAAATGCTGA 1196
Db      1384 CTCAGAGATGATGATATTTGCTCAGTTCGAGAGTGTCTCAATTTAAATGCTAA 1439

```

RESULT 11

ABV75082

ID ABV75082 standard; DNA; 1831 BP.

XX AC ABV75082;

XX DT 19-FEB-2003 (first entry)

XX DE Human beta3nT gene polynucleotide sequence.

XX KW Beta3nT; transgenic; beta-1,3-N-acetylglucosaminyltransferase;

XX OS disease system; gene; human; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 235..1429

```

FT      /*tag- a
FT      /product= "beta3GnT"
XX      W0200279413-A2.
XX      10-OCT-2002.
XX      25-MAR-2002; 2002MO-US09645.
XX      29-MAR-2001; 2001US-280706P.
XX      28-MAR-2002; 2002US-0280706.
XX      (DEL-) DELTAGEN INC.
XX      Leviten MW, Phillips R;
XX      WPI; 2003-067437/06.
XX      P-PSDB; ABB82649.
XX      New transgenic mouse comprising a disruption in a beta3GnT gene, as in
XX      vivo model to study various disease states or conditions in which
XX      beta3GnT may be implicated or involved, such as abnormal cell growth,
XX      cancer and metastasis.
XX      Disclosure; Fig 3; 55pp; English.
XX      The invention relates to a transgenic mouse comprising a disruption in a
XX      beta3GnT (beta-1,3-N-acetylglucosaminyltransferase) gene, where there is
XX      no native expression of beta3GnT gene. The transgenic mice may be used
XX      as in vivo model to study various disease states or conditions in which
XX      beta3GnT may be implicated or involved, such as abnormal cell growth,
XX      cancer and metastasis, and to evaluate various treatments or to identify
XX      agents for treating disease states or conditions, such as anxiety or
XX      depression. Animal-based disease systems may be used to identify
XX      compounds capable of ameliorating disease symptoms, as test substrates
XX      for the identification of drugs, pharmaceuticals, therapies and
XX      interventions that may be effective in treating a disease or other
XX      phenotypic characteristic of the animal. The present sequence represents
XX      a human beta3GnT polynucleotide sequence.
XX      Sequence 1831 BP; 497 A; 423 C; 443 G; 468 T; 0 other;
XX      Query Match 75.8%; Score 906.4; DB 25; Length 1831;
XX      Best Local Similarity 84.9%; Pred. No. 5,8e-270;
XX      Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
XX      1 AGATGAGTGGGGCGGCTGAAGAGTCAAGTGTCTGGGCACTCTGATGGGCAAAATGTCT 60
XX      234 AAATGAGTGTGGAGCTGCGAAGATTAAGTGTGTGGGTATCCTGTGATGGGCAAAATGTCT 293
XX      61 TCATTATTTTGAATGTGGAGTCTCCAAACACAGTAGCCCAAGACAAAATGGAAAGGGAG 120
XX      294 TCATTATTTTATATGAAGTCTCCAAAGAGCAGTACGCCAAGAAAATATGGAAGGGG 353
XX      121 GAGTAATATCCCAAGAGAGTCTTGGAGCCACCCAGCACTCCCGGGGCATCTGGA 180
XX      354 AAGTAATATATACCAAGAGAGTCTTGGAGAGATATCTACCCCTCCCGAGGCATCTGGA 413
XX      181 ACAGGGAACAGAGAGTCTGAAGTGGTGAATCCCATCTTGAACAGGGTGGCCAAATC 240
XX      414 ACCGAGACCAAGAGAGTGAACGGCAGTACACCCCATCTCTGAGCATGCTGACCAACC 473
XX      241 AGACAGGGGAGCTAGCCACATCTCCAAACACAGTACCTGAGCTATTGTGAACCAAGACT 300
XX      474 AGACGGGGGAGCGGGAGGCTCTCCAAATATAGCCCATCTGAACCTACTGCGAACCCTGACC 533
XX      301 CGACGGTCACTGACAGCTGTGACAGATTTTAATACTGCGCGGACAGATTTAAGAGCTTC 360
XX      534 TGAGGGTCACTGCGGTGGTTCAGGGTTTACAACTTCCCGGACAGATTTAAGAGCTTC 593
XX      361 TCCTGATTTGAGATCGCGGAATTAAGTCTGCTTATAGATCAACCGAAGAAATGTGCAA 420
XX      594 TCGTGTATTTGAGATCGCGCAATTAATCACTGCTTATAGATCAGCCGGATGAAGTGTGCA 653

```

```

QY      421 AGAAGCCCTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTGGCAGAGCAAG 480
XX      554 AGAAGCCCTCTTCTTGTCTGGCGATTAAAGTCCCTCACTCCACATTTTGGCAGAGCAAG 713
XX      481 CAATTCCGGGAGTCTTGGGGCCGAGAAACCAACGTAAGGAAACAGACAGTAGTAGGCTCT 540
XX      714 CAATCCGGGAATCCTGGGGCCCAAGAAAGCAACGACGAGGAACCAACGGTGGTGGAGTCT 773
XX      541 TCCTGTTGGGCAAGACACACCCCGAGAGCAACACCCCTGACCTTTCCGACATGCTTAACT 600
XX      774 TCCTGCTGGGCGAGACACCCCGAGAGCAACACCCCGACCTTTCCAGATATGCTGAAT 833
XX      601 TTGAGAGTGACAGACACCCAGAGCATCTTCATGTGGAACATATAGAGACACATCTTCAACC 660
XX      834 TTGAGAGTGAGAGACACCAAGACATCTTATGTGGAACATACAGAGACATCTTCTCAACT 893
XX      661 TGTCCCTGAAGGAGTCTGTTCTTAGTGGGTGAGCACTTCTCCTCCAGAGCGAGACT 720
XX      894 TGTCTCTGAAGGAGTCTGTTCTCAGGTGGGTAAAGTACTTCTCCCGAGACACTGAGT 953
XX      721 TTGTCTTCAAGGGCGATGATGAGTGTGTTGTGAACACCCATCACATCCTTAATCTTGA 780
XX      954 TTGTTTCAAGGGCGATGAGTGTGTTTGTGAACACCCATCACATCCTGAATTACTTGA 1013
XX      781 ATAGCTTATCCAGAGCAACCAAGACATTTGTTATAGTGAAGTGTGATGACGATCCACATGCTG 840
XX      1014 ATAGCTTATCCAGAGCAACCAAGACATCTTCTATAGGTGAAGTGAATCCACATGCTG 1073
XX      841 GGCCTCACGGGATAAAGAACTGAAGTACTACATCCAGAAAGTCTTCTACACCGGCTCT 900
XX      1074 GACCTCATCGGATTAAGAACTGAGTACTACATCCAGAAAGTGTGTTACTCTGGCTCT 1133
XX      901 ACCCACGATATCCGGGGGTGGTGGATTCCTGATCTCCGGCCCTTGGCTTGAAGCTCT 960
XX      1134 ACCCACCTATGACGGGGGAGGGGGTCTCTACTCCGGCCACCTGGCCCTGAGGCTCT 1193
XX      961 ACATGCGCATGAGCGGGTCCATCTCTACCTTATGATGATGTTTATACGGGAATGCTGC 1020
XX      1194 ACCATATCATGACCGAGTCCATCTCTACCCCATGATGACGTTTATCTGGAAATGCTCC 1253
XX      1021 TTCAGAAACTGGGCTTGTTCAGAGAGCAACAAAGCTTCAGGACATTTGATATTGAAG 1080
XX      1254 TTCAGAAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGG 1313
XX      1081 AGAAATTAAGAAATATTTGTTCTATATAGCACTTAATGTTAGTACATAGCAGAAAC 1140
XX      1314 AGAAATTAAGAAATATTTGTTCTATATAGCACTTAATGTTAGTACATAGTAAAC 1373
XX      1141 CTCAGAGATGATGATGATCTGCTCTCAGTTGCAAGTCCCTAAATTTAAATGCTGA 1196
XX      1374 CTCAGAGATGATGATGATTTGTTCTCAGTTGCAAGTGTCTCATTTAAATGCTAA 1429
XX
XX      RESULT 12
XX      AAF29255
XX      ID AAF29255 standard; cDNA; 1912 BP.
XX      AC AAF29255;
XX      DT 17-APR-2001 (first entry)
XX      DE Human beta 1,3-N-acetylglucosamine transferase G3 cDNA sequence.
XX      KW Sugar chain synthesis agent; beta 1,3-N-acetylglucosamine transferase;
XX      inflammation; cancer; metastasis; human; ss.
XX      OS Homo sapiens.
XX      PN W0200100848-A1.
XX      PD 04-JAN-2001.
XX

```

PF	29-JUN-2000; 2000W0-JP04304.
XX	
XX	29-JUN-1999; 99JP-0183437.
PR	16-MAR-2000; 2000JP-0074757.
XX	
XX	(KYOW) KYOWA HAKKO KOGYO KK.
PA	
PI	Sasaki K, Shiraishi N, Natsume A, Yamada Y, Nakagawa S, Sekine S;
XX	
XX	WPI: 2001-102895/11.
DR	P-PSDB; AAB49748.
XX	
XX	New polypeptide having beta1,3-N-acetylglucosamine transferase activity
PT	for diagnosis of inflammation, cancer and cancer metastasis, -
PT	development of remedies, and for producing glycoconjugates -
XX	
XX	Example 2; page 159-164; 195pp; Japanese.
XX	
CC	This invention relates to a sugar chain synthesizing agent that contains
CC	a polypeptide as the active ingredient, where the polypeptide has beta
CC	1,3-N-acetylglucosamine transferase activity. The polypeptides, DNA
CC	sequences encoding them and antibodies directed against the proteins are
CC	useful in the diagnosis of inflammation, cancer and its metastasis,
CC	development of remedies, and for producing sugar chains and
CC	glycoconjugates. Sequences AAB49748 - AAB49751 represent polypeptides of
CC	the invention, having beta 1,3-N-acetylglucosamine transferase activity,
CC	and polynucleotides AAF29255 - AAF29258 represent cDNA encoding them. PCR
CC	primers used in the isolation and characterisation of the cDNA sequences
CC	are represented by sequences AAF29259 - AAF29290.
XX	
XX	Sequence 1912 BP; 518 A; 430 C; 457 G; 507 T; 0 other;
XX	
Query	Match 75.8%; Score 906.4; DB 22; Length 1912;
Best	Local Similarity 84.9%; Pred. No. 6e-270;
Matches	1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
Qy	1 AGATGAGTGTGGGGCTCGAAGAGTCGAAGTCTCAAGTGTGGGTCATCTGTATGCGAAATGTCT 60
Db	232 AATGAGTGTGGAGTGTGCGAAGATAAAGTGTGGGTATCTCTGATGCGCAATGTCT 291
Qy	61 TCATTTATTGTTCGTGAAGTCTCCAAAACAGTAGCCAAAGACAAAAATGAAAGGGAG 120
Db	292 TCATTTATTTTAATGAAGTCTCCAAACAGTAGCCAAAGAAAAAATGGAAGGGG 352
Qy	121 GAGTAATAATCCGAAAGAGAAGTTCTTGGAGCCACCAGCACTCCCGGGGCATACTGGA 180
Db	352 AAGTAATAATCCAAAGAGAAGTCTTGAAGATATCTAGCCCTCCGAGGCATACTGGA 411
Qy	181 ACAGGGACAGGAGAAGCTGAACAGGTGGTAGATCCCATCTTTGAACAGGTGGCCAATC 240
Db	412 ACGGAGGACAGAGAAGCTGAACCGGAGTAGACACCCCATCTTGACATGCTGACCAACC 471
Qy	241 AGACAGGGGAGCTAGCCACATCTCCAAACAAAGTCAACCTGAGCTATTGTGAACCAACT 300
Db	472 AGACGGGGAGGGCGGCGAGGCTCTCCAATATAAGCCATCTGAACCTACTCGAACCTGACC 531
Qy	301 CGAGGCTCATGACACTGTGACAGATTTAATATCTGCCGGACAGATTTAAGACATTTC 360
Db	532 TGAGGGTCACTGCGTGGTGTACGGGTTTAAACAACCTTGGCGGACAGATTTAAGACATTTC 591
Qy	361 TCTTGTATTGTAGTCCGGGAATTACTCGCTGTCTTATAGATCAACCGAAGAAATGTGCA 420
Db	592 TGTGTATTGTAGATGCCCAATTATTCACTGCTTATAGATCAGCCGATTAAGTGTGCA 651
Qy	421 AGAAGCCCTTCITACTATTGTGGGATAAAGTCCCTCATTCACATTTTGGCAGAAGCAAG 480
Db	652 AGAAACCTTCTTGTGTGGTGGGATTAAGTCCCTCACTCCACATTTTGGCAGAAGCAAG 711
Qy	481 CAATTGGGAGTCTTGGGGCCGAGAAACCAACGTTAGGACACAGATAGTCAAGGTCT 540
Db	712 CAATCGGGAAATCTTGGGGCCAAAGAAACACGAGGGAACCAACCGGTGGTGGAGTCT 771
Qy	541 TCTGTGTGGCAAGACACCCCGACAGGACAAACCCCTGACCTTTTGGCATCTTTAAGT 600

FT /transl_except= (pos:310..315, aa:Xaa)
 FT /note= "Xaa is unknown"
 PN CA2255109-A1.
 XX 17-JUN-1999.
 PD 17-DEC-1998: 98CA-2255109.
 XX 17-DEC-1997: 97CA-2255126.
 PR (HSC-) HSC RES & DEV LP.
 PA Egan SE;
 XX WPI: 2000-148082/14.
 DR P-PSDB; AAY69698.
 XX
 PI New nucleic acids encoding a murine and human Brainiac protein, useful
 PI for detecting somatic or germline DNA lesions which are responsible for
 PT developmental syndromes or diseases including cancer.
 XX
 PS Claim 5: Page 29-30; 40pp; English.
 XX
 CC This sequence represents cDNA encoding human Brainiac protein. The
 CC complete murine Brainiac cDNA sequence (AA287185) was used to screen an
 CC EST (expressed sequence tag) database to identify human Brainiac cDNA
 CC clones, which were assembled to form the human Brainiac cDNA of the
 CC present invention. Human Brainiac has significant similarity to
 CC Drosophila melanogaster Brainiac. Drosophila Brainiac and Egghed
 CC proteins regulate adhesion between epithelial cells, this activity being
 CC dependent on the presence of Notch. Drosophila Brainiac is a secreted
 CC protein, and has sequence similarities with the Drosophila Fringe
 CC proteins. Brainiac/Egghed-mediated epithelial cell adhesion is required
 CC for cell viability, cell growth regulation and cell fate specification.
 CC Wild-type or mutant forms of mammalian Brainiac proteins may therefore be
 CC used to alter epithelial cell adhesion in a mammal. Mammalian Brainiac
 CC proteins, active fragments analogues, and nucleic acids may be used to
 CC treat diseases such as cancer, psoriasis and other skin lesions, and
 CC nervous system disorders. Mammalian Brainiac nucleic acids may also be
 CC used to detect somatic or germline DNA lesions which are responsible for
 CC developmental syndromes or diseases including cancer. The mammalian
 CC Brainiac proteins and fragments or its analogues are useful as antigens
 CC in immunoassays including enzyme-linked immunosorbent assays (ELISA),
 CC radioimmunoassays (RIA) and other non-enzyme linked antibody binding
 CC assays. Non-human transgenic animals comprising nucleotide sequences
 CC encoding human Brainiac protein (AA69698) can be used as animal models
 CC for the study of mammalian Brainiac gene function, for the screening of
 CC candidate compounds and for the evaluation of potential therapeutic
 CC interventions.
 XX
 SQ Sequence 2536 BP; 747 A; 429 C; 500 G; 803 T; 57 other;

Query Match 72.08; Score 861.6; DB 21; Length 2536;
 Best Local Similarity 83.28; Pred. No. 5.5e-256;
 Matches 973; Conservative 0; Mismatches 194; Indels 3; Gaps 1;

QY 30 TTGCTGGGCATCTCTGATGGCAATGCTCTTCAATTTATTTGTTGGAGTCTCCAA 89
 DB 28 TTGTTGGGTATCTCTGATGGCAATGCTCTTCAATTTATTTATGGAAGTCTCCAA 87
 QY 90 AACAGTAGCCAAAGACAAAATGGAAGGGAGGAGTATATATCCGAAAGAGAGTCTCG 145
 DB 88 AGCAGTAGCCAAAGAAATAATGGAAGGGGAGTATATATCCAAAGAGAGTCTCG 147
 QY 150 AAGCACCACAGCACTCCCGGGCATACTGGAACAGGGAACAGGAGAGCTGAACAGTGG 209
 DB 148 AAGATATCTACCCCTCCCGAGGCTACTGGAACCGAGAGCAAGAGAAGCTGAACCGGCAG 207
 QY 210 TACATCCCATCTTGAACAGGGTGGCCAAATCAGACAGGGGAGCTAGCCACATCTCCAAAC 269
 DB 208 TACAACCCCATCTCTGATGCTGATCCACCAACAGAGGGGGAGGGCGGCGCTCTCCAAT 267

QY 270 ARAAGTCACCTGAGCTATTGIGAACACAGACTCGACGGTCTATGACAGCTGTG---ACAGAT 326
 DB 268 ATAAGTCATCTGAACTACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTGTGTCGGGT 327
 QY 327 TTAATAATCTGCGGACAGATTAAAGACITTTCTCTTGTATTGAGATGCGGGAATTAC 386
 DB 328 TTTAAACAACCTGCGGACAGATTAAAGACTTTCTCTGTATTGAGATGCGGCAATTAT 387
 QY 387 TCGCTGCTTAAGATCAACCAAGAAATGTGCAAGAGCCCTTCTTACTATTATGGGATA 446
 DB 388 TCACCTGCTTATAGATCAGCCGGATAGTGTGCAAGAAACCTTTCTTGTGTGGCGATT 447
 QY 447 AAGTCCCTCATTCACATTTTGCAGAGCAAGCAATTCGGGAGCTTTGGGGCCGAGAA 506
 DB 448 AAGTCCCTCACTCCACATTTTGCAGAGCAAGCAATTCGGGAATCTCTGGGGCCGAGAA 507
 QY 507 ACCAAGCTAGGGACACAGACTAGTAGGCTCTTCTGTGGGCAAGACACCCCCAGAG 566
 DB 508 AGCAACGACGGGAACCAACGGTGGTGGAGTCTTCTGTGGGCAAGACACCCCCAGAG 567
 QY 567 GACAACACCCCTGACCTTTCCGACATGCTTTAAGTTTGGAGAGTGACAAGCACCAGACATC 626
 DB 568 GACAACACCCCGACCTTTCAGATATGCTGAAATTTGAGAGTCAGAAGCACCAGACATT 627
 QY 627 CTCATGTGGAACTATAGACACATTTCTCAACCTGCTCCCTGAAGCAAGTGTGTTCT 686
 DB 628 CTTATGTGGAACACTACAGACACTTTCTTCAACTTGTCTTGAAGGAAGTGTGTTCTTC 687
 QY 687 AGTGGGTGACACTTCTCTCCAGCGCAGAGATTGCTTCTTCAAGGCGCATGATGACGTG 746
 DB 688 AGTGGGTGAGTACTTCTCTCCAGCAGACATGAGTTGTTTCAAGGCGCATGACGATGT 747
 QY 747 TTTGTGAACCCCATCATCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 806
 DB 748 TTTGTGAACCCCATCATCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 807
 QY 807 GACTTGTTCATAGTGTGATCCCAATGCTGGGCTTCCAGGGGATAAGAACTGAAG 866
 DB 808 GATCTCTCATAGTGTGATGATCCCAATGCTGGGACTCATCGGGATAAGAACTGAAG 867
 QY 867 TACTACATCCAGAAAGTCTTCTACACGGGCTCTACCCACCGTATGCGGGGTGTGGGA 926
 DB 868 TACTACATCCAGAAAGTGTGTTACTTGGCTCTTACCCACCTATGCGGGGAGGGGG 927
 QY 927 TTCTGTACTCCGGCCCTTGGCTTGGAGCTGTACAGTGGGACTAGCCGGTCCATCTC 986
 DB 928 TTCTGTACTCCGGCCCTTGGCTTGGAGCTGTACCATATCACTGACCAAGTCCATCTC 987
 QY 987 TACCTATATGATGTTTATACGGGAATGTGCTTCAGAACTGGGCTTGTTCAGAG 1046
 DB 988 TACCCATGATGAGTGTATCTGGAATGTGCTTCAGAACTCGGCTCGTTCAGAG 1047
 QY 1047 AAGCACAAGGCTTCAGACATTTGATTTGAAGAGAAATAAGAAAAATATTTGTTC 1106
 DB 1048 AAACACAAGGCTTCAGACATTTGATATGAGAGAAATAAGAAAAATATATCTGCTC 1107
 QY 1107 TATATAGACCTTAATGTTAGTACATAGCAGAAACCTCAAGAGATGATTGATATCTGCT 1166
 DB 1108 TATGTAGATCTGATGTTAGTACATAGTAAACCTCAAGAGATGATTGATATTTGCT 1167
 QY 1167 CAGTTGCAAGCTCTAATTTAAATGCTGA 1196
 DB 1168 CAGTTGCAAGCTCTAATTTAAATGCTAA 1197

RESULT 15
 AAH14972
 ID AAH14972 standard; cDNA; 2442 BP.
 XX
 AC AAH14972;
 XX
 DT 26-JUN-2001 (first entry)
 XX

DE Human cDNA sequence SEQ ID NO:12897.
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
KW Homo sapiens.
OS EPI074617-A2.
PN 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 8: SEQ ID 12897; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC of an oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
SQ Sequence 2442 BP: 717 A; 433 C; 503 G; 789 T; 0 other;
Query Match 71.0%; Score 848.6; DB 22; Length: 2442;
Best Local Similarity 84.6%; Pred. No. 5.9e-252;
Matches 953; Conservative 0; Mismatches 174; Indels 0; Gaps 0;
QY 70 TGATTGTGGAGTCTCCAAAACAGTAGCCACACAAAATGAAAGGAGGAGTAATAA 129
DB 1 TTATTATGGAAGTCTCCAAAAGCAGTAGCCACACAAAATGAAAGGAGGAGTAATAA 60
QY 130 TCCCGAAGAGAGTCTTGGAGGCCACCCAGCACTCCCGGGCATACTGGAAACAGGGAAC 189
DB 61 TACCCAAAGAGAGTCTTGGAGGATATCTACCCCTCCCGAGGATCTACTGGAAACAGGAGCC 120
QY 190 AGGAGAGCTGAACAGGTGGTACAAATCCCATCTTGAACAGGTGGCCATCAGACAGGGG 249
DB 121 AAGAGAGCTGAACCGGCGAGTACAACCCCATCTCTGAGCATGCTTGACCAACAGGAGGGG 180

QY 250 AGCTAGCCACATCTCCAAACACAAAGTCACCTGAGCTATTGTGAACACAGACTGACGGTCA 309
DB 181 AGCGGGCAGGCTCTCCAATATAAGCCATCTGAACCTACTGCGAAGCTGACCTGAGGGTCA 240
QY 310 TGACAGCTGTGACAGATTTTAAATCTCCGCGACAGATTTAAAGACTTCTCTGTGAT 369
DB 241 CGTCGGTGGTACGGGTTTTAACAACTTCCGCGACAGATTTAAAGACTTCTCTGTGAT 300
QY 370 TGAGATGCCGGAATTAATCTCGCTCTTATAGATCAACCGAAGAAATGTGCAAGAGCCCT 429
DB 301 TGAGATGCCGGAATTTATCACTGCTTATAGATCAGCCGATTAAGTGTGCAAGAGAACTT 360
QY 430 TCTTACTATTGGCGATAAAGTCCCTCAATCCACATTTTCCGAGAGGAGCAAGCAATTCGG 489
DB 361 TCTGTGTTGCTGGCGATTAAGTCCCTCACTCCACATTTTCCGAGAGGAGCAAGCAATTCGG 420
QY 490 AGTCTTGGGCGCGAGAAACCAAGTAGGGAACACAGACAGTGTGAGGGTCTTCTCTGTGG 549
DB 421 AATCTTGGGCGCGAGAAACCAAGTAGGGAACCAAGTGTGAGGGTCTTCTCTGTGG 480
QY 550 GCAAGACACCCCGAGAGGAGCAACCCCTGACCTTTCGAGACATGCTTAAGTTGAGAGTG 609
DB 481 GCCAGACACCCCGAGAGGAGCAACCCCTGACCTTTCAGATATGCTGAAATTTGAGAGTG 540
QY 610 ACAAGCACAGGACATCCTCATGTGCACTATAGACACACATTTCTCAACCTGTCCCTGA 669
DB 541 AGAAGCACAGGACATTTATGTGGAACCTACAGACACACTTCTCAACTTGTCTCTGA 600
QY 670 AGAAGTGTGTTTCTTAGTGGGTGAGCACTTCTCTGTCGACGACGAGATTTGTCTTCA 729
DB 601 AGAAGTGTGTTTCTTAGTGGGTGAGCACTTCTCTGTCGACGACGAGATTTGTCTTCA 660
QY 730 AGGCGATGATGAGTGTGTTGTGAACACCCATCCTTAATTAAGTTGAGAGTGT 789
DB 661 AGGCGATGAGTGTGTTGTGAACACCCATCCTGAATTAAGTTGAGAGTGT 720
QY 790 CCAAGACAAAGCAAGAGCTTCTCATAGTGACGTGATCCACAAATGCTGGGCTCACC 849
DB 721 CCAAGACAAAGCAAGAGCTTCTCATAGTGATGTGATCCACAAATGCTGGGCTCACC 780
QY 850 GGGATAAGAACTGAAGTACTACATCCAGAGCTCTTCTACACCCGCGCTTACCCACCT 909
DB 781 GGGATAAGAACTGAAGTACTACATCCAGAGTGTGTTTACTCTGGCCTTACCCACCT 840
QY 910 ATGCGGGGTGGTGGATTCCTTACTCCGGCCCCCTTGCCTTGGGCTGTACAGTGGA 969
DB 841 ATGCGGGGTGGTGGATTCCTTACTCCGGCCCCCTTGCCTTGGGCTGTACAGTATCA 900
QY 970 CTAGCGGGTCCATCTCTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 1029
DB 901 CTAGCGGGTCCATCTCTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 960
QY 1030 TGGGCTTGTTCAGAGAGCAACAAAGGCTTACAGACATTTGATATTTGAAGAGAAATA 1089
DB 961 TGGGCTTGTTCAGAGAGCAACAAAGGCTTACAGACATTTGATATTCAGAGAGAAATA 1020
QY 1090 AGAAATATTTGTTCTTATATAGACCTAATGTACTACATACGACAGAAACCTCAGAGA 1149
DB 1021 AAAATACATCTGCTCTCTATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
QY 1150 TGATTGATATCTGCTCTCAGTTTCAAAGCTCTTAATTTAAATGCTGA 1196
DB 1081 TGATTGATATTTGCTCTCAGTTTCAAAGCTCTTAATTTAAATGCTGA 1127

Search completed: October 20, 2003, 14:45:55
Job time : 343.33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 14:27:23 : Search time 2702.64 Seconds
(without alignments)
10755.472 Million cell updates/sec

Title: US-09-804-357b-1
Perfect score: 1196
Sequence: 1 agatgagtgtgggcgtcga.....gtccfaatttaaatgctga 1196

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match: 0%
Maximum Match: 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estm:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_estl:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pin:*

20: em_gss_vit:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gssl:*

29: gb_gss2:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	542	45.3	587	10	BE623172 up49a10.x
2	533.4	44.6	547	14	CD560301 B0410D02-
3	529	44.2	589	9	AW495241 UI-M-BH3-
4	521.8	43.6	832	9	AU137528 AU137528

5	521	43.6	592	10	BF467111	BF467111 UI-M-CG0p
6	517	43.2	517	9	AA119132	AA119132 mp66f04.r
7	517	43.2	743	14	CB554366	CB554366 MMSP0063-
8	499.6	41.8	1201	9	AL553132	AL553132 AL553132-
9	488	40.8	505	9	AA204363	AA204363 mu30e11.r
10	483	40.4	692	10	BG106243	BG106243 602289389
11	479.2	40.1	609	10	BE100018	BE100018 UI-R-RJ1-
12	474	39.6	876	13	BX370083	BX370083 BX370083
13	466.2	39.0	488	10	BE632125	BE632125 uu57d09.y
14	456.8	38.2	476	9	AA986646	AA986646 uc82d09.y
15	456.4	38.2	1053	12	BM557603	BM557603 AGENCOURT
16	455.4	38.1	457	9	AA733556	AA733556 vu71f09.r
17	441	36.9	501	10	BE685687	BE685687 uu57d09.x
18	440.2	36.8	664	14	BY724171	BY724171 BY724171
19	440	36.8	489	10	BG062994	BG062994 H3002H04-
20	440	36.8	686	14	BY754595	BY754595 BY754595
21	433	36.2	655	14	BY753386	BY753386 BY753386
22	432.2	36.1	659	10	BG741357	BG741357 602631909
23	417	34.9	543	9	AW530344	AW530344 UI-R-BU0-
24	416.4	34.8	477	10	BE944633	BE944633 UI-M-BH3-
25	412.8	34.5	558	12	BM537435	BM537435 ha92c09.g
26	408.4	34.1	410	10	BF469505	BF469505 UI-M-BH3-
27	403.6	33.7	680	14	CB457941	CB457941 716060 MA
28	401.4	33.6	964	13	B0893660	B0893660 AGENCOURT
29	400	33.4	490	9	AI408668	AI408668 EST326959
30	388.8	32.5	945	13	B0369620	B0369620 603596654
31	379.6	31.7	610	10	B0619419	B0619419 BB619419
32	378.2	31.6	484	9	AI407754	AI407754 EST236044
33	375.4	31.4	392	9	AA929654	AA929654 vy75a02.r
34	371	31.0	430	10	BE946424	BE946424 UI-M-BH3-
35	369.8	30.9	593	10	AW958548	AW958548 EST370618
36	369	30.9	632	10	BG721682	BG721682 602698930
37	363.8	30.4	865	13	B0427178	B0427178 603955596
38	361	30.2	719	12	B1157193	B1157193 602921987
39	360	30.1	372	9	AA209930	AA209930 mu40a10.r
40	360	30.1	373	10	BE631971	BE631971 uu56b02.y
41	359	30.0	701	9	AV724042	AV724042 AV724042
42	344.4	28.8	405	9	AW456055	AW456055 UI-M-BH3-
43	337	28.2	444	9	AI070816	AI070816 UI-R-C2-m
44	334	27.9	383	9	AI112158	AI112158 UI-R-YO-m
45	333.8	27.9	813	12	BI664280	BI664280 603289034

ALIGNMENTS

RESULT 1
BE623172
LOCUS
DEFINITION
up49a10.x1 Soares_mouse.MMGB.bccll Mus musculus cDNA clone
IMAGE:2749434 3' Similar to TR:054904 054904 UDP-GAL:BETAGLCNAC
BETA 1.3-GALACTOSYLTRANSFERASE-1. ; mRNA sequence.
ACCESSION
BE623172
VERSION
BE623172.1
KEYWORDS
GI:9903578
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 587)
AUTHORS
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL
Tumor Gene Index
COMMENT
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaps@email.nih.gov
This clone is available royalty-free through LILNL; contact the
IMAGE Consortium (info@image.lilnl.gov) for further information.
MGI:1035806
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 452.
Location/Qualifiers
1..587


```

Db      121  |||||AGTCCCTCAATCCACATTTTGGCCAGAGGCAAGCAATTCGGAGTCTTGGGGCCGAGAA 180
Qy      507  ACCAAGCTAGGAGACAGACAGTAGTGAGGGTCTTCTGCTGGCAAGACACCCGCCAGAG 566
Db      181  ACCAAGCTAGGAGACAGACAGTAGTGAGGGTCTTCTGCTGGCAAGACACCCGCCAGAG 240
Qy      567  GACAACACCCCTGACCTTTCCGGACATGCTTAAGTTTGAGAGTGACAAGCACCAGACATC 626
Db      241  GACAACACCCCTGACCTTTCCGGACATGCTTAAGTTTGAGAGTGACAAGCACCAGACATC 300
Qy      627  CTCATGTGGAACTATAGACACATCTTCAACCTGTGCTCCCTGAGGAAGTGTGTTCTT 686
Db      301  CTCATGTGGAACTATAGACACATCTTCAACCTGTGCTCCCTGAGGAAGTGTGTTCTT 360
Qy      687  AGTGGGTGAGCACTTCTCTGTCAGAGCAGAGTTTGTCTTCAAGGGCGATGACGCTG 746
Db      361  AGTGGGTGAGCACTTCTCTGTCAGAGCAGAGTTTGTCTTCAAGGGCGATGACGCTG 420
Qy      747  TTGTGAGACCCATCACATCTTATTACTTGAATAGCTTATCCAGAGCAAGGCCAAA 806
Db      421  TTGTGAGACCCATCACATCTTATTACTTGAATAGCTTATCCAGAGCAAGGCCAAA 480
Qy      807  GACTTGTTCATAGGTGACGTGATCCACATGCTGGCCCTCACCGGGGATAGAACTGAAG 866
Db      481  GACTTGTTCATAGGTGACGTGATCCACATGCTGGCCCTCACCGGGGATAGAACTGAAG 540
Qy      867  TACTACA 873
Db      541  TACTACA 547

```

RESULT 3

```

AW495241 589 bp mRNA linear EST 24-1FB-2000
LOCUS UT-M-BH3-aus-b-04-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
DEFINITION UT-M-BH3-aus-b-04-0-UI 3', mRNA sequence.

```

```

ACCESSION AW495241
VERSION AW495241.1 GI:7065522

```

```

KEYWORDS EST.
SOURCE Mus musculus (house mouse)

```

```

ORGANISM Mus musculus

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 589)

```

```

BONALDO,M.F., Lennon,G. and Soares,M.B.

```

```

Normalization and subtraction: two approaches to facilitate gene
discovery

```

```

Genome Res. 6 (9), 791-806 (1996);

```

```

97044477
8889548

```

```

Contact: Chin, H

```

```

National Institute of Mental Health

```

```

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

```

```

20892-9643, USA

```

```

Tel: 301 443 1706

```

```

Fax: 301 443 9890

```

```

Email: mEST@mail.nih.gov

```

```

Oligo-dT track not found, Not 1 site shown in beginning of sequence

```

```

is likely internal to the message. cDNA library Preparation: M.B.

```

```

Soares Lab Clone distribution: Researchers may obtain BMAP cDNA

```

```

clones from RESEARCH GENETICS. It should be noted that Bento Soares

```

```

is generating a small number of additional specialized

```

```

non-redundant arrays of BMAP cDNAs whose availability will be

```

```

considered under appropriate and limited collaborative arrangements

```

```

Seq primer: M13 Forward

```

```

POLYA-No. Location/Qualifiers

```

```

1. 589

```

```

/organism="Mus musculus"

```

```

/mol_type="mRNA"

```

```

/strain="C57BL/6J"

```

```

/db_xref="taxon:10090"

```

FEATURES

```

source

```

```

/clone="UI-M-BH3-aus-b-04-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NIH_BMAP_M_S4"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldio, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=hippocampus
TAG_SEQ=TTCGA*

```

```

BASE COUNT 174 a 136 c 154 g 124 t 1 others
ORIGIN

```

```

Query Match 44.2%; Score 529; DB 9; Length 589;
Best Local Similarity 99.8%; Pred. No. 4.6e-127;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGATGAGTGTGGGGCTCGAAGAGTCAAGTTGCTGGGCATCCTCGATGATGGCAAAATGTCT 60
Db 60 AGATGAGTGTGGGGCTCGAAGAGTCAAGTTGCTGGGCATCCTCGATGATGGCAAAATGTCT 119
Qy 61 TCATTTATTGATGCTGAAGTCTCCAAAACAGTACCAAGACAAAATGGAAGGGAG 120
Db 120 TCATTTATTGATGCTGAAGTCTCCAAAACAGTACCAAGACAAAATGGAAGGGAG 179
Qy 121 CAGTAATAATCCGAAAGAGAAGTTCTGGAAGCCACCCAGCACTCCCGGGGCATCTGGA 180
Db 180 CAGTAATAATCCGAAAGAGAAGTTCTGGAAGCCACCCAGCACTCCCGGGGCATCTGGA 239
Qy 181 ACAGGGAACAGGAGAAGCTGAACAGTGGTACAAATCCCATCTTGAACAGGGTGGCCATC 240
Db 240 ACAGGGAACAGGAGAAGCTGAACAGTGGTACAAATCCCATCTTGAACAGGGTGGCCATC 299
Qy 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAAGTCACTGAGCTATGTGAACCCAGACT 300
Db 300 AGACAGGGGAGCTAGCCACATCTCCAAACACAAGTCACTGAGCTATGTGAACCCAGACT 359
Qy 301 CGAGGGTCATGACAGCTGTGACAGATTTTAAATCTGCCGGACAGATTTAAAGACTTTC 360
Db 360 CGAGGGTCATGACAGCTGTGACAGATTTTAAATCTGCCGGACAGATTTAAAGACTTTC 419
Qy 361 TCTTGTATTGAGATGGCGGAAATCTGCTGCTTATAGATCAACCGAAGAAATGTGCAA 420
Db 420 TCTTGTATTGAGATGGCGGAAATCTGCTGCTTATAGATCAACCGAAGAAATGTGCAA 479
Qy 421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTTCCACATTTTGGCAGAGGCAAG 480
Db 480 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTTCCACATTTTGGCAGAGGCAAG 539
Qy 481 CAATTGGGGAGTCTTGGGGCGGAGAAACCAACAGTAGGGAACCCAGACAGTA 530

```

Db 540 CAATTCGGGAGTCTTGGGCGGAGAACCAAGTANGGAACACAGACAGTA 589

RESULT 4
AUI37528
LOCUS AUI37528 PLACE1 Homo sapiens cDNA clone PLACE1006678 5', mRNA
DEFINITION AUI37528
ACCESSION AUI37528
VERSION AUI37528.1 GI:10998067
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 832)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isigai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Location/Qualifiers
1..832
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1006678"
/tissue_type="placenta"
/clone_lib="PLACE1"
/notes="Vector: pME18SFL3"
BASE COUNT 248 a 193 c 194 g 194 t 3 others

Query Match 43.6%; Score 521.8; DB 9; Length 832;
Best Local Similarity 80.9%; Pred. No. 3.9e-125;
Matches 656; Conservative 0; Mismatches 149; Indels 6; Gaps 4;
QY 77 GGAAGTCTCCAAACACAGTAGCCACAGACACAAAATGGAAGGGAGGAGTAATAATCCCGNA 136
Db 9 GAAAGTCTCCAAAGCAGTAGTACCAAGAAAATGGAAGGGAGGAGTAATAATACCCAA 68
QY AGAGAAGTCTCGAAGCCACCCAGCAGTCTCCCGGGCATCTGGAACAGGAGGAGGAGAA 196
Db 69 AGAGAAGTCTCGAAGTATCTACCCCTCCGAGGATATCTGGAACAGGAGGAGGAGAA 128
QY GCTGAACAGGTGGTACAAATCCCAATCTTGACAGGGTGGCCAAATCAGACAGGAGGAGTAGC 256
Db 129 GCTGAACCGGCGAGTACAAACCCCATCTGAGCATGCTGAGCAACAGGAGGAGGAGGCGG 188
QY 257 CACATCTCCAAACACAGTACCTGAGCTATTGTAACAGAGTCTGACCGGTCATGACACC 316
Db 189 CAGGCTCTCAATATAAGCCATCTGAACTACTCTGAACTGAGTCTGAGGTCACGTCGGT 249
QY 317 TGTGACAGATTTTAAATATCTGCGGACAGATTAAAGACTTTCTCTCTATTATGAGATG 376
Db 249 GGTACGGGTTTAAACAACTTGGCGGACAGATTAAAGACTTTCTCTCTATTATGAGATG 308
QY 377 CCGGAATTAATCGCTGCTTATAGATCAACCGAAGAAATGTGCAAGAACCCCTTCTTACT 436
Db 309 CCGCAATTAATCACTGCTTATAGATCAACCGGAGTAAGTGTGCAAGAACCCCTTCTTGT 368

QY 437 ATTGGCGATAAAGTCCCTCATTTCCACATTTTCCAGAGGCAAGCAATTCGGAGTCTTG 496
Db 369 GCTGGCGATTAAAGTCCCTCATTTCCACATTTTCCAGAGGCAAGCAATTCGGAGTCTTG 428
QY 497 GGGCCGAGAAACCAAGTAGGGAACGACAGACAGTACTAGGGTCTCTCTGTTGGGCAAGC 556
Db 429 GGGCCGAGAAACCAAGTAGGGAACGACAGGAAACCAAGGTTGGGAGTCTCTCTGTTGGGCAAGC 488
QY 557 ACCCCAGAGGAGCAACACACCTGACCTTTCCGACATGCTTTAAGTTTGAAGTGAAGCAAGCA 616
Db 489 ACCCCAGAGGAGCAACACACCTGACCTTTCCGACATGCTTTAAGTTTGAAGTGAAGCAAGCA 548
QY 517 CCAGGACATCTCATGTGGAACATATAGACACATCTTCAACCTGTCCTCAAGCAAGT 676
Db 549 CCAGGACATCTCATGTGGAACATATAGACACATCTTCAACCTGTCCTCAAGCAAGT 608
QY 677 GCTGTTCTTCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 736
Db 609 GCTGTTCTTCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 668
QY 737 TGATGAGGTGTTGTGACACCATCATCATCTTAACTT--GAATAGCTTATCCAG 794
Db 569 TGACGATGTTTGTGACACCATCATCATCTTAACTT--GAATAGCTTATCCAG 728
QY 795 AGCAAGCCAAAGACTT--GTTTATAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 851
Db 729 ACCAAAGCCAAAGACTT--GTTTATAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 788
QY 852 GATAAGAACTGAAG-TACTACATCCCAAA 881
Db 789 GATAAGAACTGAAGTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 819
RESULT 5
LOCUS BF467111
DEFINITION UI-M-CG0p-bre-g-04-0-UI-s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
ACCESSION BF467111
VERSION BF467111.1 GI:11536294
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 592)
AUTHORS Bonaudo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
Oligo-dT track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA-No. Location/Qualifiers
1..592
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"


```

|||||
157 GACAGCAGGACATCCCTCATGTGGAACATATAGAGACACATTTCTCAACCTGTCCCTG 98
Db

669 AAGGAAGTGTCTTCTTAGTGGGTGAGCACTTCCTGTCCAGAGCAGAGTTTGTCTTC 728
Qy
|||||
97 AAGGAAGTGTCTTCTTAGTGGGTGAGCACTTCCTGTCCAGAGCAGAGTTTGTCTTC 38
Db

729 AAGGGCGATGATGACGTGTTTGTGAACACCCATCACA 765
Qy
|||||
37 AAGGGCGATGATGACGTGTTTGTGAACACCCATCACA 1
Db

RESULT 7
CB554366
LOCUS CB554366 743 bp mRNA linear EST 01-JUN-2003
DEFINITION MMSP0063_H02 MMSP Macaca mulatta cdna, mRNA sequence.
ACCESSION CB554366
VERSION CB554366.1 GI:31303561
KEYWORDS EST.
SOURCE Macaca mulatta (rhesus monkey)
ORGANISM Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
1 (bases 1 to 743)
Katze,M.G., Bungarner,R., Korth,M., Feldman,R., Amjadi,M. and
Holzman,T.
Expressed sequence tags from Rhesus macaque spleen
Contact: Holzman T
Katze Lab
University of Washington
Box 358070, Seattle, WA 98105-8070, USA
Tel: 206 732 6156
Fax: 206 732 6055
Email: ted@locke.hs.washington.edu
Similar to GenBank entry BC030579 BC030579 Homo sapiens,
UDP-GlucNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 1,
clone MGC:26071 IMAGE:4828158, mRNA, complete cds. 5/2002
Plate: MMSP0063 row: H column: 02.
FEATURES
Location/Qualifiers
1..743
/organism="Macaca mulatta"
/mol_type="mRNA"
/db_xref="taxon:9544"
/sex="male"
/cell_type="monocytes"
/dev stage="adult"
/clone.lib="MMSP"
/note="Organ: spleen"
BASE COUNT 200 a 182 c 175 g 182 t
ORIGIN
Query Match 43.2% Score 517; DB 14; Length 743;
Best Local Similarity 82.9% Pred. No. 6,6e-124;
Matches 614; Conservative 0; Mismatches 125; Indels 2; Gaps 2;
Qy 185 GGACAGAGAGAGTGTGAACAGGTGTGTACATCCATCTTGAACAGGGTGGCCATCAGAC 244
Db 1 GGCACAGATGATGAACAGGGCAGTACACACCCCATCTGAGCATGTAGCCAAACCCAGAC 60
Qy 245 AGGGGAGTACCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACCACTCGAC 304
Db 61 GGGGAGGGGGGAGGGTCTCTCAATATAGCCATCTGTGAATGTGAGCGCTGACCTGAG 120
Qy 305 GGTCTATGACAGTGTGACAGATTTTAATATCTGCCGACAGATTTAAAGACATTTCTT 364
Db 121 GGTCAATCATGTTGTACAGTTTAAATACCTTGGCGGACAGATTTAAAGACATTTCTGT 180
Qy 365 GTATTGAGATGCCGGGAATTACTCGTGTCTTATAGATCAACCCAGAAATATGCAAGAA 424
Db 181 GTATCTGATGCGCGCAATTATTCACTGCTTATAGATCAACCCGGAATAGTGTGCAAGAA 240

```

```

425 GCCCTTCTTACTATTCGGGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAGCAAT 484
Db
241 ACCCTTCTTGTGCTGGCGATTAAAGTCCCTCACTCCACATTTTGCAGAGGCAAGCAAT 300
Qy
485 TCGGGAGTCTTGGGCGGAGAAACCAACCTAGGGAACCAAGACAGTACTGAGGCTTCTCT 544
Db
301 TCGGGAATCTTGGGCGGAGAAAGCAACACGCGGGAACCAACCCGCTTTCAGATATGCT 360
Qy
545 GTTGGGCAAGACACACCCCGAGAGGAAACCAACCCCTGACCTTTCGGACATGCTTAAGTT 604
Db
361 GCTGGGCGAGACACCCCGAGAGGAAACCAACCCCGACCTTTCAGATATGCTGAATTTGA 420
Qy
605 GAGTGACAGCAGCAGGACATCTCATGTGGAACTATAGAGACACATTCCTCAACCTGTC 664
Db
421 GAGTGACAGCAGCAGCAGCAGATCTTATGTGGAATATACAGAGACATTTCTTCAACTGTC 480
Qy
665 CTTGAAGGAGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 724
Db
481 TCTGAAGGAGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 540
Qy
725 CTTCAAGGCGGATGATGACGTGTTTGTGAACACCCCATCACATCTTAATTACATAG 784
Db
541 TTTCAAGGCGGATGATGATGTTTGTGAACACCCCATCACATCTTGAATTTACTTGAATAG 600
Qy
785 CTTATCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 844
Db
601 TTTATCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 660
Qy
845 TCACCGGGGATGAAGAACTGAAGTACTACATCCAGAAAGTCTTCTTACACCGGGGTACCC 904
Db
661 TCATCGGGATGAAGAACTGAAGTACTACATCCAGAAAGT-TGTTACTCTGGCCTCTA-CC 718
Qy
905 ACCGATATCGCGGGGTGGTGG 925
Db
719 ACCATATGACAGGGGGGGGG 739

RESULT 8
LOCUS AL553132 1201 bp mRNA linear EST 31-MAY-2003
DEFINITION AL553132 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA
clone CS001074YK14 5-PRIME, mRNA sequence.
ACCESSION AL553132
VERSION AL553132.2 GI:31274946
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
L.I.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12892683.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 1713.f For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS001074BF07QPLcluster=1713.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS001074BF07QPL.
FEATURES
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS001074YK14"
/tissue_type="PLACENTA COT 25-NORMALIZED"

```

```
/clone_lib="HOMO sapiens PLACENTA COI 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      302 a      280 c      331 g      251 t      37 others
ORIGIN

Query Match      41.8%; Score 499.6; DB 9; Length 1201;
Best Local Similarity 82.3%; Pred. No. 2.6e-119;
Matches 604; Conservative 4; Mismatches 123; Indels 3; Gaps 3;

Qy 1 AGATGACTGTGGCGCTCGAAGAGTCAAGTTCGTGGGCATCTCGATGATGGCAATGTCT 60
Db 312 AAATGAGTGTGGACGTCGAAGAATAAAGTGTGGGTATCTCGATGATGGCAATGTCT 371
Qy 61 TCATTTATTTGATGTGGAGAGTCTCCAAAAGAGTAGCCAAAGACAAAATGGAAGGAG 120
Db 372 TCATTTATTTATGTAAGTCTCCAAAGACAGTAGCCAAAGAAATGGAAGGGG 431
Qy 121 GAGTAATAATCCCGAAAGAGAAGTCTGGAAGCCACCCAGCACTCCCGGGCATACTGGA 180
Db 432 AAGTAATAATACCCAAAGAGAAGTCTGGAAGATATCTACCCCTCCCGAGGCATACTGGA 491
Qy 181 ACAGGGAACAGGAGAACTCAACAGGTGGTACATCCATCTTGAACAGGTCGCCAATC 240
Db 492 ACCGAGAGCAGAGAGCTGAACCGGAGTACACCCCATCTGAGCATGCTGACCACAC 551
Qy 241 AGACGGGGAGCTAGCCATCTCCAAACAAAGTCACCTGAGCTATTGTGAACACGACT 300
Db 552 AGACGGGGAGCGGGCGAGCTCTCCAAATATAGCCATCTGACTACTGCGAACCIGACC 611
Qy 301 CGACGGTCATGACAGCTGTGACAGATTTTAATATCTGCGGGACAGATTTAAAGATTTTC 360
Db 612 TGAGGGTCACCTGGTGGTTACGGGTTTAAACAACTTGGCGGACAGATTTAAAGACTTC 671
Qy 361 TCTTGATTTGAGATCGCGGAATTACTCGCTGCTTATAGATCAACCGAAGAAATGTGCAA 420
Db 672 TGCTGTAATTTGAGATCGCGCAATTAATCTGCTTATAGATCAGCGGATTAAGTGTGCAA 731
Qy 421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATCTCCACATTTTGCCAGAAGCAAG 480
Db 732 AGAAACCTTTCTTCTGCTGGCATTAAGTCCCTCACTCCACATTTTGCCAGAAGCAAG 791
Qy 481 CAATTCGGAGTCTTGGGCGGAGAACCAACGCTAGGAAACCAAGACAGTAGTCAGGCTCT 540
Db 792 CAATTCGGGAATCTTGGGCGCAAGAAGC-AMCAGGGAAACCAACGGTGGTGGAGTCT 850
Qy 541 TCCGTGTTGGGCAAGACACCCCGAGAGGACACCAACCCCTGACCTTTCCGACATCCTTAAGT 600
Db 851 TCCTGCTGGGCGACACACCCCGAGAGGACACCAACCCCGACCTTTTCAGATATGCTGAAT 910
Qy 601 TTGAGAGTGACACACCCAGGACATCTCATGTGGAACTATAGAGACATTTCTCAACC 660
Db 911 TTGAGAGTGAGACACCAAGACATCTTATGTGGGANTACAGAGACACTTTCTCAACT 970
Qy 661 TGTCCCTGAAGGAAG-TGCTGTTTCTTAGTGGGTGAGCACTTCCTCTCCAGACGACAG 719
Db 971 TGTCTCTGAGGAAGTGTCTGTTTCTCAGGTGGGTAAATACTT-CTGCCAGACACTGAG 1029
Qy 720 TTTGCTCTCAAGGG 733
Db 1030 TTTGTTTCAAGGS 1043

RESULT 9
AA204363/c LOCUS
DEFINITION mu30ell.r1 Soares.thymus_2NbMT Mus musculus cDNA clone IMAGE:640940
5' similar to TR:G1150971 G1150971 NEUROGENIC SECRETED SIGNALING
PROTEIN. ;, mRNA sequence.
ACCESSION AA204363
VERSION AA204363.1 GI:1801007
```

```
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 505)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:392932
Possible reversed clone: similarity on wrong strand
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 495.
FEATURES
Location/Qualifiers
1..505
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:640940"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares_thymus_2NbMT"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Bonaudo."
BASE COUNT 120 a 120 c 121 g 144 t
ORIGIN

Query Match 40.8%; Score 488; DB 9; Length 505;
Best Local Similarity 99.8%; Pred. No. 2.3e-116;
Matches 499; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 381 AATTACTCCTGCTTATAGATCAACCGAAGAAATGTGCAAGAAGCCCTTCTTACTATTG 440
Db 505 AATTACTCCTGCTTATAGATCAACCGAAGAAATGTGCAAGAAGCCCTTCTTACTATTG 446
Qy 441 GCGATAAAGTCCTCATTCACATTTTCCAGAGGCAAGCAATTCGGGAGTCTTTGGGCG 500
Db 445 GCGATAAAGTCCTCATTCACATTTTCCAGAGGCAAGCAATTCGGGAGTCTTTGGGCG 386
Qy 501 CGAGAAACCAAGCTAGGNAACCCAGACAGTAGTAGGGTCTTCCTCTTGGGCAAGACACC 560
Db 385 CGAGAAACCAAGCTAGGNAACCCAGACAGTAGTAGGGTCTTCCTCTTGGGCAAGACACC 326
Qy 561 CCAGAGGACCAACCCCTGACCTTTCCGACATGCTTAAGTTTGAGAGTGACAAAGCACAG 620
Db 325 CCAGAGGACCAACCCCTGACCTTTCCGACATGCTTAAGTTTGAGAGTGACAAAGCACAG 266
Qy 621 GACATCCTCATGTGGAACCTATAGAGACACATTTCTTCAACCTGCTCCCTGAAGGAAGTGTG 680
Db 1030 TTTGTTTCAAGGS 1043
```


Db	265	GACATCTCTCTGTGGAACTATAGAGACACAATCTTCAACCTGTCCCTGAAGGAATGCTGT	206
QY	681	TTTCTTAGTGGGTGAGCACTTCTCTGTCCAGACGACAGATTTGTCTTCAAGGGCGAIGAT	740
Db	205	TTTCTTAGTGGGTGAGCACTTCTCTGTCCAGACGACAGATTTGTCTTCAAGGGCGAIGAT	146
QY	741	GACGTGTTGTGACACACCCATCACATCCTTAAATTAATTCTTGAATAGCTTATCCCAAGAGCAAA	800
Db	145	GACGTGTTGTGACACACCCATCACATCCTTAAATTAATTCTTGAATAGCTTATCCCAAGAGCAAA	86
QY	801	GCCTAAGACACTTGTTCATAGGTGACGTGATCCCAATGCTGGGCGCTCAC -CGGGA-AAAGAA	859
Db	85	GCCAAGACACTTGTTCATAGGTGACGTGATCCCAATGCTGGGCGCTCAC -CGGGA-AAAGAA	26
QY	860	ACTGAAGTACTACATCCACAG	879
Db	25	ACTGAAGTACTACATCCACAG	6
RESULT 10			
LOCUS	BG106243	692 bp	linear
DEFINITION	602289589F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4384414	5'	EST 30-JAN-2001
ACCESSION	BG106243		
VERSION	BG106243.1	GI:12600089	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 692)		
JOURNAL	NIH-MGC http://mgc.ncbi.nlm.nih.gov/		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgbpbs@mail.nih.gov		
	Tissue Procurement: Louis Staudt, M.D., Ph.D.		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LLAM10064 row: a column: 23		
	High quality sequence stop: 687.		
FEATURES			
source	1..692		
	Location/Qualifiers		
	organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4384414"		
	/tissue_type="lymphoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/clone_lib="NIH_MGC_85"		
	/note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;		
	Site_2: SalI; Cloned unidirectionally; oligo-dr primed.		
	Average insert size 1.867 Kb. Library enriched for		
	full-length clones and constructed by Life Technologies.		
	Note: this is a NIH_MGC Library."		
BASE COUNT	186 a 166 c 167 g 172 t		1 others
ORIGIN			
Query Match	40.4%	Score 483;	DB 10; Length 692;
Best Local Similarity	84.5%	Pred. No. 4.9e-115;	
Matches	566; Conservative	0; Mismatches 100;	Indels 3; Gaps 2
Qy	265	CRAACACAAAGTCACCTGACCTATTGTGAACACAGACTCGACGGGTGTCATCACAGCTGTGCACAG	324
Db	23	CCATATAGGCATCTGAACTACTCGGAACCTGACCTGAGGGTCACTCGGTGGTTACGG	82
Qy	325	ATTTTTAAATTCGCGGACAGATTAAAGACTTTCTTCTTATTTGAGATGCCGAATT	384

```

Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-No.      Location/Qualifiers
1. .609
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/lab_host="PH10B (Life Technologies)"
/clone_lib="UI-R-BJ1"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
library is a subtracted library derived from the following
tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.
For a detailed description of the library from which this
clone was derived, please visit our web site at
ratest.eng.uiowa.edu. The subtraction has been previously
described in (Bonafide, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_SEQ=None found"
BASE COUNT      169 a 139 c 165 g 134 t 2 others
ORIGIN
Query Match      40.1%; Score 479.2; DB 10; Length 609;
Best Local Similarity 91.8%; Pred. No. 4.7e-114;
Matches 505; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 1 AGATGAGTGTGGGCGTCGAAGAGTCAAGTTCGTGGCGATCTGTGATGGCAATGTCT 60
Db 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 60 AATGAGTGTGGGCGTCGAGAGTAAAGTTCGTGGCGATCTGTGATGGCAATGTCT 119
Db 60 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 TCATTATTGATGTGGAGTCTCCAAACAAAGTAGCCCAAGACAAACAAATGGAAGGGAG 120
Db 61 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 120 TCATCTATTGATGTGGAGTCTCCAAACAAAGTAGCCCAAGACAAAGATGGAAGGGGG 179
Db 120 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 GACTATAATACCGAAGAGAGTCTGGAAGCCACCCAGCACTCCCGGGGATATGGA 180
Db 121 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 180 GAGTGATAATACCGAAGAGAGTCTGGAAGCCACCCAGCACTCCCGGGGATATGGA 239
Db 180 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 181 ACAGGGAACAGGAGAGTCTGAACAGTGTGTACAAATCCCATCTTGAACAGGTGGCCAAATC 240
Db 181 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 240 ATAGGGAACAGGAGAGTCTGAACAGTGTGTACAAATCCCATCTTGAACAGGTGGCCAAATC 299
Db 240 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 241 AGACAGGGAGCTAGCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACT 300
Db 241 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 300 AGACGGGGATCTGTTACATCTTCCAAACACAGTCACTGAGCTATTGTGAACAGACT 359
Db 300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 301 CGACGCTCATGACAGTGTGACAGATTTTAATTAATCTGCGGACAGATTTAAGACTTTC 360
Db 301 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 360 CGACAGTCAATGACAGTGTGACAGATTTTAATTAATCTGCGGACAGATTTAAGACTTTC 419
Db 360 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 361 TCTGTATTGATGATCGCGAATTACTCGGTGCTTATAGATCAACCGAAGAAATGTGCA 420
Db 361 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 420 TCTGTATTGATGATCGCGAATTACTCGGTGCTTATAGATCAACCGAAGAAATGTGCA 479
Db 420 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 421 AGAAGCCCTTCTTACATATTGGCGATTAAGTCCCTCATTCACATTTTGCCAGAGGCAAG 480
Db 421 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 480 AGAAGCCCTTCTTACATATTGGCGATTAAGTCCCTCATTCACATTTTGCCAGAGGCAAG 539
Db 480 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 481 CAATTCCGGAGCTCTTGGGCGGAGAAACCAACCTAGGAAACACAGACTAGTAGAGGTCT 540
Db 481 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 540 CAATTCCGGAGCTCTTGGGCGGAGAAACCAACCTAGGAAACACAGACTAGTAGAGGTCT 599
Db 540 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 541 TCCTGTGGG 550
Db 541 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 600 TCCTTTTGGG 609
Db 600 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
FEATURES
source
LOCUS BX370083 876 bp mRNA linear EST 08-MAY-2003
DEFINITION BX370083 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0DI074YK14 5-PRIME, mRNA sequence.
ACCESSION BX370083
VERSION BX370083.1 GI:30449879
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 876)
AUTHORS Li W.B., Gruber C., Jessee J., and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 1713.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAG036ZC11_CS03443_1&cluster=1713.f.
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID: CS0BAG036ZC11_CS03443_1.
Location/Qualifiers
1. 876
/organism="Homo sapiens"
/db_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI074YK14"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/notes="First strand cDNA was primed with a NotI-oligo (3T)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 248 a 172 c 196 g 258 t 2 others
ORIGIN
Query Match      39.6%; Score 474; DB 13; Length 876;
Best Local Similarity 86.7%; Pred. No. 1.2e-112;
Matches 534; Conservative 0; Mismatches 80; Indels 2; Gaps 1;
QY 583 TTTCGGACATGCTTAAGTTTTCAGAGTGACAGACACAGG--ACATCTCATGTGGAACCTA 640
Db 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 3 TTGAGGATGCTGCTGAATTTTGAGAGTGAGAGACCAAGNNACATTTCTATGTGGAACCTA 62
Db 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 641 TAGAGACACATCTTCAACCTGTCCTCAAGGAAGTGTCTTCTAGTGGGTGAGCAC 700
Db 641 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 63 CAGAGACACATTTCTTCAACTGTCTCTCAAGGAAGTGTCTTCTAGTGGGTGAGTAC 122
Db 63 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 701 TTCTGTCCAGACGACAGATTTTGTCTTCAAGGGGATGATGACGTGTTTGTGAACACCCA 760
Db 701 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 123 TTCTGTCCAGACACTGAGTTTGTTCAGGGCGATGACCATGTTTGTGTGAACACCCA 182
Db 123 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 761 TCACATCTCTTAATTAATTAAGTCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGG 820
Db 761 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 183 TCACATCTCTGAATTAATTAAGTCTTATCCAGACCAAGCCAAAGACTCTCTTCATAGG 242
Db 183 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 821 TCACGTGATCCACAATGCTGGCCCTCACCGGATGAAGAACTGAAGTACTACATCCACGA 880
Db 821 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 243 TGATGTGATCCACAATGCTGGACCTCATCGGGATGAAGAACTGAAGTACTACATCCACGA 302
Db 243 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 881 AGTCTTCTACACCGCGCTACCCACCGTATGCGGGGGTGTGATTCCTGTACTCCGG 940
Db 881 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 303 AGTTGTTTACTCTGGCCCTTACCCACCTCTACCCAGGGGAGGGGGTTCCTTACTCCGG 362
Db 303 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 941 CCCCCCTTGGCCTTGAGGCTGTACAGTCGCACTAGCCGGGTCCCATCTCTACCTATTGATGA 1000
Db 941 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 363 CCACCTGGCCCTTGAGGCTGTACCATATCACTACACAGGTCCATCTCTACCCCATGTGA 422
Db 363 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

QY 1001 TCTTTATACGGGAATGTCCTTCAGAAACTGGGCTTGTCTCCAGAGACGACAAAGGCTT 1060
Db 423 CGTTTATACGGAATGTCCTTCAGAAACTGGGCTTGTCTCCAGAGACGACAAAGGCTT 482
QY 1061 CAGGACATTTGATATTGAAGACAAAATAAAGAAAATATTTGTCCTATATAGACCTAAT 1120
Db 483 CAGGACATTTGATATTGAAGACAAAATAAAGAAAATATTTGTCCTATATAGACCTAAT 542
QY 1121 GTTAGTACATACGACAAAACCTCAAGAGATGATTCATATCTGGTCTCAGTTGCAAAAGTCC 1180
Db 543 GTTAGTACATACGACAAAACCTCAAGAGATGATTCATATCTGGTCTCAGTTGCAAGTCC 602
QY 1181 TAAATTAAATGCTGA 1196
Db 603 TCAITTAATGCTAA 618

RESULT 13
BE632125/c
LOCUS
DEFINITION
us57d09.y1 Soares_thymus_2NDMT Mus musculus cDNA clone
IMAGE:3376049 5' similar to TR:Q54904 Q54904 UDP-GAL:BETAGLCNAC
BETA 1.3-GALACTOSYLTRANSFERASE-1. ;, mRNA sequence.
ACCESSION
BE632125
VERSION
BE632125.1 GI:9914895
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 488)
REFERENCE
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1085653
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 455.
FEATURES
source
1..488
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3376049"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares_thymus_2NDMT"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TG-TACCAATCTGAAGTCGAGCGCGGCTTTTATTTTATTTTATTTTATTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 136 a 112 c 108 g 132 t
ORIGIN
Query Match 39.0%; Score 466.2; DB 10; Length 488;
Best Local Similarity 99.4%; Pred. No. 1.1e-110;
Matches 468; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 669 AAGGAAGTCTGTTCTTAGTGGGTGAGCACTTCTCTCCAGACGACAGATTGTCCTTC 728
Db 488 AAGGAAGTCTGTTCTTAGTGGGTGAGCACTTCTCTCCAGACGACAGATTGTCCTTC 429
QY 729 AAGGGCGATGATGACGTGTTGTGAACACCCATCATCTTAATTAATTAATAGCTTA 788
Db 428 AAGGGCGATGATGACGTGTTGTGAACACCCATCATCTTAATTAATTAATAGCTTA 369
QY 789 TCCAAGAGCAAGCCAAAGACTTGTTCATAGTGCATGATCCCAATAGCTGGGCTCAC 848
Db 368 TCCAAGAGCAAGCCAAAGACTTGTTCATAGTGCATGATCCCAATAGCTGGGCTCAC 309
QY 849 CGGGATAAGAACTGAAGTACTACATCCAGAGAGTCTTCTACACGGCGTCTACCCACG 908
Db 308 CGGGATAAGAACTGAAGTACTACATCCAGAGAGTCTTCTACACGGCGTCTACCCACG 249
QY 909 TATGCCGGGGGTGGTGCATTCTGTACTCCGGCCCCCTTGCCTTGAGGCTGTACAGTGG 968
Db 248 TATGCCGGGGGTGGTGCATTCTGTACTCCGGCCCCCTTGCCTTGAGGCTGTACAGTGG 189
QY 969 ACTAGCCGGGTCCATCTCTACCTTATGATGATGTTTATACGGGAATGTCCTTCAGAAA 1028
Db 188 ACTAGCCGGGTCCATCTCTACCTTATGATGATGTTTATACGGGAATGTCCTTCAGAAA 129
QY 1029 CTGGGCTTGTCCAGAGAGACAAAGGCTTCAGGACATTTGATATTGAAGAGAAAAT 1088
Db 128 CTGGGCTTGTCCAGAGAGACAAAGGCTTCAGGACATTTGATATTGAAGAGAAAAT 69
QY 1089 AAGAAAATATTTGTTCCCTATATAGACCTTAATGTTAGTACATAGCAGAAAA 1139
Db 68 AAGAAAATATTTGTTCCCTATATAGACCTTAATGTTAGTACATAGCAGAAAA 18

RESULT 14
AA986646 476 bp mRNA linear EST 28-MAY-1998
LOCUS
uc82d09.y1 Sugano mouse kidney mklia Mus musculus cDNA clone
IMAGE:1432145 5' similar to TR:Q24157 Q24157 NEUROGENIC SECRETED
SIGNALING PROTEIN. ;, mRNA sequence.
ACCESSION
AA986646
VERSION
AA986646.1 GI:3168300
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 476)
REFERENCE
Marra.M., Hillier.L., Allen.M., Bowles.M., Dietrich.N., Dubuque.T.,
Geisel.S., Kucaba.T., Lacy.M., Le.M., Martin.J., Morris.M.,
Schellenberg.K., Steptoe.M., Tan.F., Underwood.K., Moore.B.,
Theising.B., Wylie.T., Lennox.G., Soares.B., Wilson.R. and
Waterston.R.
The WashU-HMI Mouse EST Project
Unpublished
Contact: Marra M/Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:916213
Seq primer: primer name ambiguous
High quality sequence stop: 474.
FEATURES
source
1..476
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1432145"

```

Contact: Robert Strausberg, Ph.D. Email: cgaabs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM12754 row: c column: 21 High quality sequence stop: 547.	
FEATURES	
source	
1..1053	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="taxon:9606"	
/clone="IMAGE:5739980"	
/tissue_type="duodenal adenocarcinoma, cell line"	
/lab_host="DH10B (phage-resistant)"	
/clone_lib="NIH_MGC_88"	
/note="Organ: small intestine; Vector: pCMV-SPORT5; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."	
BASE COUNT	288 a 240 c 219 g 306 t
ORIGIN	
Query Match	38.2%; Score 456.4; DB 12; Length 1053;
Best Local Similarity	87.5%; Pred. No. 4.9e-108;
Matches	499; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY	627 CTCATGTGGAACCTATAGAGACACATCTTCAACCTGTCCTCGAGGAAGTGTCTTCTT 686
Db	1 CTTATGTGGAACCTATAGAGACACATCTTCTCAACTTGTCTCTGAAGGAAGTGTCTTC 60
QY	687 AGTGGGGTGGACACTTCCTGTCCAGACCGAGAGTGTGCTTCAAGGGGATGATGACGTG 746
Db	61 AGTGGGGTGGACACTTCCTGTCCAGACCGAGTGTGTTTTCGAAGGGCGGATGACGATGT 120
QY	747 TTTGTGAACACCCATCACATCCTTAAATTACTTGAATAGCTTATCCAAAGACCAAGCCAAA 806
Db	121 TTTGTGAACACCCATCACATCCTTGAATTACTTGAATAGTATTATCCAAAGACCAAGCCAAA 180
QY	807 GACTTGTTCATAGGTGAGTGTATCCCAAACTGCTGGGCTCACCGGGATAGAACTGAAG 866
Db	181 GATCTCTTCATAGGTGATGTGATCCCAATGCTGGACCTCATCGGGATAGAACTGAAG 240
QY	867 TACTACATCCAGAAAGTCTTCTACACCGGGCTCTACCCACCGTATCCGGGGTGGTGA 926
Db	241 TACTACATCCAGAAAGTGTGTTACTCTGGCCTCTACCCACCTATGCGAGGGGAGGGGG 300
QY	927 TTCCTGTACTCCGGCCCTTGGCTTGGAGGCTGTACAGTGGGAGTACGCCGGTCCATCTC 986
Db	301 TTCCTGTACTCCGGCCACCTGGCCCTGAGGCTGTACCATATCACTGACCAAGGTCCATCTC 360
QY	987 TACCTATTGATGATGTTTATACGGGAATGTCCTTCAGAACTGGGCTTGTTCACAGAG 1046
Db	361 TACCCCATTTGATGACGTTTATCTGGAATGTGCTTCAGAAACTCGGCCCTCTCCAGAG 420
QY	1047 AAGCACAAGGCTTCAGGACATTTGATATTGAAGAGAAATAAGAAAAATATTTGTCTCC 1106
Db	421 AAACACAAGGCTTCAGGACATTTGATATCGAGGAGAAACAAANAATATCATCTGTCTCC 480
QY	1107 TATATAGACCTTAATGTTAGTACATAGCAAAAAACCTCAAGAGATGATGATATCTGGTCT 1166
Db	481 TATGTAGATCTGATGTTAGTACATAGTAGAAAACCTCAAGAGATGATGATGATTTGGTCT 540
QY	1167 CAGTTGCAAGTCTCTAAATTTAAATGCTGA 1196
Db	541 CAGTTGCAAGTCTCTAAATTTAAATGCTGA 570

Search completed: October 20, 2003, 19:37:12
Job time : 2719.64 secs

```
Db 418 ACNTGGGNGM-----NGTNGNGNGTGGGCGNMGNGNMGNCARYTNAARYTNGTN 468
QY 552 AAGACACCCCGAGGACACACCCCTGACCTTTCGGACATGCTTAAGTTGAGATGAC 611
Db 469 TTYTNYTNGNGTNGGNGNSGNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 528
QY 612 AAGCAGGAGACATCTCATATGGAACATATAGACACACATCTTCAACCTGCTCCCTGAAG 671
Db 529 GARTYTGAYATHVTCNARTCGGAYTYTACNGARGAYTYTYAAYTACNYTNAAR 588
QY 672 GAAGTCTGTTCTTAGTGGGTGAGCACTTCTCTGTCAGACCCAGAGTTTGTCTCAAG 731
Db 589 GARYTTCAYTNGCARGNTGGTNGTNGCNGCTGYCCNCCNCCNCCNCCNCCNCCNCCNCCN 648
QY 732 GCGGATCATGACGTGTGTGGAACACCCATCATCTTAAITACTTGAATACCTTAACC 791
Db 649 GNGAYGAYGAYGNTYGTNCAYGTNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 703
QY 792 AAGAGCAAGCCAAAGACATGTTTCATAGGTGAGCTGATCCACAAATGCTGGGCTCACCGG 851
Db 704 -GGAYCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 762
QY 852 GATAAGAACTGAAGTACTACATCCAGAGTCTTCTACA---CCGGGCTCTACCCACCG 908
Db 763 AATACNARGTNAARTATTTAATCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 822
QY 909 TATCGGGGGGTGGTGGTCTGCTACTCGGCCCCCTTGCCTTGAGGTGTACAGTGGG 968
Db 823 TAYCGNGGNGGNGGNGTAYGTNATGWSNMHNGNCCNCCNCCNCCNCCNCCNCCNCCNCCN 882
QY 969 ACTAGCGGGTCTATCTACCTTACCTTATGATGTTTATACGGGATGTCCTTCAGAAA 1028
Db 883 ATGARGAYGNGARVNTYTNWSNATGAYGAYGTNTTYGNGNATGTYTNGMNGN 942
QY 1029 CTGGGCTTGTTCAGAGAGCAAGAGCTTCAGGACATTTGATATT--GAAGAGAAA 1085
Db 943 YTNNGNYTNWSNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 1002
QY 1086 AATAAGAAAAATTTGTCTCTATATAGACCTAATCTTAGTACATAGCAGAAAACTCAA 1145
Db 1003 GAYCCNYTNAGCCNTGYTNTAYTMGNGNYTNTYNTCTNCCNCCNCCNCCNCCNCCNCCN 1062
QY 1146 GAGATGATGATATCTGGTCTAGTTGCAAGTCCCTAATTTAAATGTC 1193
Db 1063 GARATGTGACNATGTGGGCGNYTNGTNGACNGAYGARGGNYTNAARTGY 1110
```

RESULT 2

```
US-09-459-133-3
: Sequence 3, Application US/09459133
: Patent No. 6416988
: GENERAL INFORMATION:
: APPLICANT: Conklin, Darrell C.
: APPLICANT: Yamamoto, Gayle
: APPLICANT: Jaspers, Stephen R.
: APPLICANT: Gao, Zeren
: TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
: FILE REFERENCE: 98-77
: CURRENT APPLICATION NUMBER: US/09/459,133
: PRIOR FILING DATE: 1999-12-10
: PRIOR FILING DATE: 1998-12-10
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 3
: LENGTH: 1191
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Degenerate sequence
: FEATURE:
: NAME/KEY: misc_feature
```

```
: LOCATION: (1)...(1191)
: OTHER INFORMATION: n - A,T,C or G
US-09-459-133-3
```

```
Query Match 13.6%; Score 163; DB 4; Length 1191;
Best Local Similarity 34.0%; Pred. No. 8.7e-43;
Matches 261; Conservative 110; Mismatches 379; Indels 18; Gaps 2;
```

```
QY 426 CCCTTCTTACTATTGGCGATAAAGTCCCTCATCTTCACATTTTGGCAGAGCAAGCAATT 485
Db 442 CCAATAYTNTYNTGCGTNAARWSNARCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 501
QY 486 CGGGAGTCTTGGGCGGAGAAACCAACCTAGGGAACAGACAGTAGTGAGGTCTTCTCG 545
Db 502 MNGARACNTGGGWNCCNCC-----NCCNGNATHMGNVNTYNTTY 546
QY 546 TTGGGCAAGACACCCCGAGAGGACCAACCCCTGACCTTTCGGACATCTTCAACCTGTC 605
Db 547 YTNNTGNGNWSNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 606
QY 506 AGTGAAGACAGCAGCAGACATCTTCATGTGGAACTATAGACACATCTTCAACCTGTC 665
Db 607 WSNMNGNTAYWSNGAYTNTYNTGCGAYTNTYNTGAGYTGTCNCCNTTAAAYCARACN 666
QY 666 CTGAAGAGAGTCTGTTTCTTAGGTGGTGAGCACTTCTCTCCAGAGCAGAGTTTCTC 725
Db 667 YTNAAARGAYTNTYNTYNTGCTGTTGAGCAACCCATCATCTTAAITACTTGAATGAC 726
QY 726 TTCAAGGGGATGATGAGTGTCTGTAACCCATCATCTTAAITACTTGAATGAC 785
Db 727 YTNMNGNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 786
QY 786 TTATCAAGAGCAAGCAAGCAAGCTTGTTCATAGTACGCTGATCCACAATGCTGGGCT 845
Db 787 YTNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 846
QY 846 CACCGGGATGAAGAACTGAAGTACTACATCCAGAGTCTTCTACACCGGCTTACCCA 905
Db 847 YTNMNAARCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 906
QY 906 CGGTATGCGGGGTGGTGGTCTCTGCTACCTGCTACCGGCCCCCTGCTTGGGCTGACAGT 965
Db 907 GNTAYGNCNWSGNGGNGGNTAYGTNATHGCGNMGNYTNGCNCNTTGGYTYNTNNGN 966
QY 966 GCGATAGCGGGTCCATCTCTACCTATGATGATTTATACGGGATGTCCTTCAG 1025
Db 967 GCGNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 1026
QY 1026 AAACCTAGCGGCTTGTTCAGAGAGCAAGGCTTCAGACATTTGATTAAGAGAGAAA 1085
Db 1027 GCNTYNGGNYTNGTNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 1086
QY 1086 AATAAGAAAAATTTGTCTCTAGTTCAGTTCGAAAGTCTTAAITTTAAATGTC 1193
Db 1087 ACNCCNGAYCA---YTGCGNTTYMGNAAAYTNTYNTGNTMNGNCCNCCNCCNCCNCCNCCN 1143
QY 1146 GAGATGATGATATCTGCTCTAGTTCGAAAGTCTTAAITTTAAATGTC 1193
Db 1144 GCNWSNATHMGNVNTYNTGAGAACARYTNCARGAYCCNMGNYTNCARTGY 1191
```

RESULT 3

```
US-09-459-133-14.
: Sequence 14, Application US/09459133
: Patent No. 6416988
: GENERAL INFORMATION:
: APPLICANT: Conklin, Darrell C.
: APPLICANT: Yamamoto, Gayle
: APPLICANT: Jaspers, Stephen R.
: APPLICANT: Gao, Zeren
: TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
: FILE REFERENCE: 98-77
: CURRENT APPLICATION NUMBER: US/09/459,133
```



```
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07

; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match      3.7%   Score 44.6;   DB 4;   Length 2095;
Best Local Similarity 46.0%;   Pred. No. 0.00054;
Matches 189;   Conservative 0;   Mismatches 219;   Indels 3;   Gaps 1;

QY 383 TTACTCGCTGCTTATAGATCAACCGAAGAAATGTGCAAGAGCAATTCGGGAGTCTTTGGGCGG 442
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 510 TCACCTTCACACTTCGAGAGCATTCAAACCTGCTCTCATCAAAATCCATTTCTGGTCAICT 569
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 443 GATAAAGTCCTCATTCACATTTTGCCAGAAGCAAGCAATTCGGGAGTCTTTGGGCGG 502
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 570 GGTGACCTCCACCCCTTCAGATGTGAAAGCCAGCGAGGCCATTAGAGTTACTTTGGGGTGA 529
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 503 AGAAACCAACGTAGGGAACCAAGACAGTAGTGAGGGTCTTCTCTGTGGGCAAGACACGCC 562
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 630 AAAAAGTCTTGGTGGGATATGAGTTCTTACATTTTCTTATTAGGCCAAGAGGCTGA 689
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 563 AGAGGACAACCCCTGACCTTCGGACATGCTTAAGTTTGAGAGTGACAAAGCAGGGA 622
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 690 AAGGAGAGACAAATGTTGGCATTGTCC---TTAGAGGATGAACACCTTTCTTTATGGTGA 746
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 623 CATCTCATGTGGAACTATAGAGACACATTTTCAACCTGTCCCTGAAGGAAGTGTGT 682
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 747 CATATCGACAAGATTTTITAGACACATATATAACCTGACCTTGAACACCATTAATGGC 806
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 683 TCTTAGTGGGTGAGCACCTTCTGTCAGACGAGAGTTTGTCTTCAAGGGCGAIGATGA 742
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 807 ATTCAGGTGGTAACTAGTGTTCGCCCAATGCCAAGTAGCTAATGAAGACAGACACTGA 866
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 743 CGTGTCTTGAACACCCATCACATCCTTAATTACTTGAATAGCTTATCCAA 793
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 867 TGTTTCATCAATACTAGCAATTTAGTGAAGTATCTTTTAAACCTAAACCA 917
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 8
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOXLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMNU
```

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)883-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match          3.6%; Score 43.2; DB 1; Length 7218;
Best Local Similarity 7.1%; Pred. No. 0.0033;
Matches 24; Conservative 172; Mismatches 140; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGCGTCGAAGAGTCAAGTTGCTGGGCATCTCTGATGCGCAAAATGTCT 60
Db 1352 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1293
QY 61 TCATTATTGTTGTTGGAAGTCTCAAAAACAGTAGCCCAAGACAAAAATGGAAGGGAG 120
Db 1292 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1233
QY 121 GAGTAATAATCCGGAAGAGAACTTCTGGAAGCCACCCAGCACTCCCGCGGCATCTGCA 180
Db 1232 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1173
QY 181 ACAGGGAACAGGAGAGAGTGAACAGTGGTAGCAATCCCATCTTGAACAGGGTGGCAATC 240
Db 1172 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1113
QY 241 AGACAGGGAGCTAGCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACT 300
Db 1112 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1053
QY 301 CGACGGTCACTGACAGCTGTGACAGATTTTAATAATC 336
Db 1052 CGACCTGCACCAAGTCGGAATTAATCTGTGAGC 1017

RESULT 9
US-09-055-097-2
Sequence 2, Application US/09055097
Patent No. 5955282
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESS: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/055,097
FILING DATE: Filed Herewith
CLASSIFICATION:
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
```

```
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Cerrone, Michael C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0490 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1434 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: DUODNOT02
CLONE: 1705085
US-09-055-097-2

Query Match          3.5%; Score 41.8; DB 2; Length 1434;
Best Local Similarity 48.8%; Pred. No. 0.0035;
Matches 143; Conservative 0; Mismatches 147; Indels 3; Gaps 1;

QY 474 AGGCAAGCAATTCGGGAGTCTTGGGCGGAGAAACCAACGTAGGAAACCCAGACAGTGTG 533
Db 298 AGAAAGCCATTCGGGCTTCGTGGGGGGGCTGCCGAGGCCCGGGGCTCAGGGTACAG 357
QY 534 AGGGTCTTCCTGTTGGGGAAGACACCCCAAGGAGAAC---CACCTGACCTTTTCGGAC 590
Db 358 AGCTATTCTTGTGGAGAGCCGAAACGACAGCACCCCGTGTGGGTTCACAGGGGAGT 417
QY 591 ATGCTTAAGTTTGAGAGTGACAAAGCACCAGGACATCCTCATGTGGAACATATAGACACA 650
Db 418 GACCTGGGCTCGAGTACAGCAGCCCGAGGGGAGATCTTGCAGGGCGGCTTCCAGGACTCC 477
QY 651 TTCTTCAACCTGTCCTCTGAAGGAAGTGTCTTTCTTAGTGTGGGTGAGCAGCTTCCTGTCCA 710
Db 478 TACCGCAACCTCACCTTAAGACCTCAGCGGGCTGAACCTGGCTGAGAAACACTGCCGCC 537
QY 711 GACGAGAGTTTGTCTTCAAGGGCGCATGATGACGTTGTTTGAACACCCATCA 763
Db 538 ATGGCCCGATAGTCTCTCAAGACGAGCATGATGTATGTATGTAAGTCCCTGA 590

RESULT 10
PCT-US93-06748-1
Sequence 1, Application PC/TUS9306748
GENERAL INFORMATION:
APPLICANT: Yung-Kang Chow
APPLICANT: Martin S. Hirsch
APPLICANT: Debra P. Meiril
APPLICANT: Joan C. Kaplan
APPLICANT: Richard T. D'Aquila
TITLE OF INVENTION: CONVERGENT COMBINATION ANTI-VIRAL
TITLE OF INVENTION: THERAPY
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06748
FILING DATE: 19930719
CLASSIFICATION:
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
```


Query Match 2.9%; Score 35.2; DB 3; Length 2601;
Best Local Similarity 50.0%; Pred. No. 0.73; Mismatches 88; Indels 0; Gaps 0;
Matches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 657 AACCTGTCCTGAAGCAAGTGTCTTTCTTAGTGGGTGAGCACTTCCTGTCCAGAGCCA 716
DB 1609 AAGCACTAACAGAGTAATACCACTAACAGAGCAAGCTAGAACTGGCAGANAACA 1668
QY 717 GAGTTTGTCTTCAAGCGCGATGATGACGTTTGTGAACACCCATCACATCCTTAATAC 776
DB 1669 GAGAGATTCTAAAAGAACCAAGTACATGAGTGTATATGACCCATCAAAAGACTTAATAG 1728
QY 777 TTGAATAGCTTATCCAGAGCAAGCAAGCAAGTGTCTTATAGGTGACGTGATCCA 832
DB 1729 CAGAAATACAGAGCAAGCGGCAAGGCCAATGGACATATCAAAATTTATCAAGAGCCA 1784

RESULT 14
US-09-117-217-13
; Sequence 13, Application US/09117217
; Patent No. 6221578
; GENERAL INFORMATION:
; APPLICANT: de BETHUNE, Marie-Pierre
; APPLICANT: HERTOGS, Kurt
; APPLICANT: PAUWELS, Rudi
; TITLE OF INVENTION: METHOD OF MANAGING THE CHEMOTHERAPY OF PATIENTS WHO ARE
; TITLE OF INVENTION: HIV POSITIVE BASED ON THE PHENOTYPIC DRUG SENSITIVITY
; FILE REFERENCE: 1377-125P
; CURRENT APPLICATION NUMBER: US/09/117,217
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 2601
; TYPE: DNA
; ORGANISM: HIV-HXB2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (750)..(2435)
; OTHER INFORMATION: Reverse Transcriptase
US-09-117-217-13

Query Match 2.9%; Score 35.2; DB 3; Length 2601;
Best Local Similarity 50.0%; Pred. No. 0.73; Mismatches 88; Indels 0; Gaps 0;
Matches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 657 AACCTGTCCTGAAGCAAGTGTCTTTCTTAGTGGGTGAGCACTTCCTGTCCAGAGCCA 716
DB 1609 AAGCACTAACAGAGTAATACCACTAACAGAGCAAGCTAGAACTGGCAGANAACA 1668
QY 717 GAGTTTGTCTTCAAGCGCGATGATGACGTTTGTGAACACCCATCACATCCTTAATAC 776
DB 1669 GAGAGATTCTAAAAGAACCAAGTACATGAGTGTATATGACCCATCAAAAGACTTAATAG 1728
QY 777 TTGAATAGCTTATCCAGAGCAAGCAAGCAAGTGTCTTATAGGTGACGTGATCCA 832
DB 1729 CAGAAATACAGAGCAAGCGGCAAGGCCAATGGACATATCAAAATTTATCAAGAGCCA 1784

RESULT 15
US-09-735-487-7
; Sequence 7, Application US/09735487
; Patent No. 6528251
; GENERAL INFORMATION:
; APPLICANT: de BETHUNE, Marie-Pierre
; APPLICANT: HERTOGS, Kurt
; APPLICANT: PAUWELS, Rudi
; TITLE OF INVENTION: METHOD OF MANAGING THE CHEMOTHERAPY OF PATIENTS WHO ARE
; TITLE OF INVENTION: HIV POSITIVE BASED ON THE PHENOTYPIC DRUG SENSITIVITY
; TITLE OF INVENTION: OF HUMAN HIV STRAINS
; FILE REFERENCE: 1377-125P

; CURRENT APPLICATION NUMBER: US/09/735,487
; CURRENT FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 09/117,217
; PRIOR FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2601
; TYPE: DNA
; ORGANISM: HIV-HXB2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(492)
; OTHER INFORMATION: gag Polyprotein
US-09-735-487-7

Query Match 2.9%; Score 35.2; DB 4; Length 2601;
Best Local Similarity 50.0%; Pred. No. 0.73; Mismatches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
Matches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 657 AACCTGTCCTGAAGCAAGTGTCTTTCTTAGTGGGTGAGCACTTCCTGTCCAGAGCCA 716
DB 1609 AAGCACTAACAGAGTAATACCACTAACAGAGCAAGCTAGAACTGGCAGANAACA 1668
QY 717 GAGTTTGTCTTCAAGCGCGATGATGACGTTTGTGAACACCCATCACATCCTTAATAC 776
DB 1669 GAGAGATTCTAAAAGAACCAAGTACATGAGTGTATATGACCCATCAAAAGACTTAATAG 1728
QY 777 TTGAATAGCTTATCCAGAGCAAGCAAGCAAGTGTCTTATAGGTGACGTGATCCA 832
DB 1729 CAGAAATACAGAGCAAGCGGCAAGGCCAATGGACATATCAAAATTTATCAAGAGCCA 1784

Search completed: October 20, 2003, 19:40:24
Job time : 83.2177 secs


```
121 GAGTAATAATCCCGAAGAGAGTTCTGAGAGCCACCAGCACATCCCGCGGCATACTGGA 180
181 ACAGGGAACAGAGAGAGCTGAACAGGTGGTCAATCCCATCTTGAACAGGGTGGCCAAATC 240
181 ACAGGGAACAGAGAGAGCTGAACAGGTGGTCAATCCCATCTTGAACAGGGTGGCCAAATC 240
241 ACACAGGGGAGTACGCCATCTCCAAACACAAAGTCACTGAGCTATTGTGAACAGACT 300
241 ACACAGGGGAGTACGCCATCTCCAAACACAAAGTCACTGAGCTATTGTGAACAGACT 300
301 CCAGGTCATGACAGCTGTGACAGATTTTAATAATCTGCCGACACAGATTTAAAGACTTTC 360
301 CCAGGTCATGACAGCTGTGACAGATTTTAATAATCTGCCGACACAGATTTAAAGACTTTC 360
361 TCTTGTTATTGAGATGCCGGAAATTAATCTGCTGCTTATAGATCAACCGAAGAAATGTGCA 420
361 TCTTGTTATTGAGATGCCGGAAATTAATCTGCTGCTTATAGATCAACCGAAGAAATGTGCA 420
421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 480
421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 480
481 CAATTGGGAGTCTTGGGCGGAGAAACCAAGTAGGGAACAGACAGTAGTGAAGGTCT 540
481 CAATTGGGAGTCTTGGGCGGAGAAACCAAGTAGGGAACAGACAGTAGTGAAGGTCT 540
541 TCCCTGTGGCAAGACACCCAGAGGACACACCCCTGACTTTCGGACATGCTTAAGT 600
541 TCCCTGTGGCAAGACACCCAGAGGACACACCCCTGACTTTCGGACATGCTTAAGT 600
601 TTGAGAGTGACAAAGCACCAGGACATCTCTCATGTGGAATATAGAGACACATTTCTCAACC 660
601 TTGAGAGTGACAAAGCACCAGGACATCTCTCATGTGGAATATAGAGACACATTTCTCAACC 660
661 TGTCCCTGAAGAGTGTCTTCTTAGTGGGTGAGCACTTCCTGTCCAGACGAGAGT 720
661 TGTCCCTGAAGAGTGTCTTCTTAGTGGGTGAGCACTTCCTGTCCAGACGAGAGT 720
721 TTGCTCTCAAGGGCGATGATGAGTGTCTTGAACACCCATCACATCTTAATTACTTGA 780
721 TTGCTCTCAAGGGCGATGATGAGTGTCTTGAACACCCATCACATCTTAATTACTTGA 780
781 ATAGCTTTATCCAGAGCAAGCAAGACTTGTTCATAGGTGACGTGATCCACAAATGCTG 840
781 ATAGCTTTATCCAGAGCAAGCAAGACTTGTTCATAGGTGACGTGATCCACAAATGCTG 840
841 GGCCTCACCGGATGAAGAACTGAAGTACTACATCCAGAAAGTCTTCTACACCGGCTCT 900
841 GGCCTCACCGGATGAAGAACTGAAGTACTACATCCAGAAAGTCTTCTACACCGGCTCT 900
901 ACCACCGTATGCCGGGGTGGTGAATCTCTACTCCGGCCCTTCCCTTGAGGCTGT 960
901 ACCACCGTATGCCGGGGTGGTGAATCTCTACTCCGGCCCTTCCCTTGAGGCTGT 960
961 ACAGTCCGACTAGCCGGGTCCATCTCTACCCCTATTGATGATGTTTATACGGGAATGTGCC 1020
961 ACAGTCCGACTAGCCGGGTCCATCTCTACCCCTATTGATGATGTTTATACGGGAATGTGCC 1020
1021 TTCAGAACTGGCCCTTGTTCAGAGAGCAAAAGGCTTCAGACATTTGATTTGAAG 1080
1021 TTCAGAACTGGCCCTTGTTCAGAGAGCAAAAGGCTTCAGACATTTGATTTGAAG 1080
1081 AGAAAAATGAAGAAAAATTTCTTCTATATAGACTTAATGTTTACTACATACGAAAAAC 1140
1081 AGAAAAATGAAGAAAAATTTCTTCTATATAGACTTAATGTTTACTACATACGAAAAAC 1140
1141 CTCAGAGATGATGATATCTGGTCTCAGTTCGAAAGTCTTAATTTAAATGCTGA 1196
1141 CTCAGAGATGATGATATCTGGTCTCAGTTCGAAAGTCTTAATTTAAATGCTGA 1196
```

```
US-09-804-006-1
; Sequence 1, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/126001
; CURRENT APPLICATION NUMBER: US/09/804,006
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/292,228
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1196
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-804-006-1
```

```
Query Match 100.0%; Score 1196; DB 10; Length 1196;
Best Local Similarity 100.0%; Pred No 0;
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AGATGAGTGTGGGCGTCGAAGAGTCAAGTTGCTGGGCATCCTGATGATGGCAATGTCT 60
Db 1 AGATGAGTGTGGGCGTCGAAGAGTCAAGTTGCTGGGCATCCTGATGATGGCAATGTCT 60
Qy 61 TCATTTATTTGATGTGGAAGTCTCCAAAACAGTAGCCCAAGACAAATATGGAAGGGAG 120
Db 61 TCATTTATTTGATGTGGAAGTCTCCAAAACAGTAGCCCAAGACAAATATGGAAGGGAG 120
Qy 121 GAGTAATAATCCCGAAGAGAGTTCTGGAAGCCACCAGCACTCCCGGGCATACTGGA 180
Db 121 GAGTAATAATCCCGAAGAGAGTTCTGGAAGCCACCAGCACTCCCGGGCATACTGGA 180
Qy 181 ACAGGGAACAGAGAGAGCTGAACAGGTGGTGAATCCCATCTTGAACAGGGTGGCCAAATC 240
Db 181 ACAGGGAACAGAGAGAGCTGAACAGGTGGTGAATCCCATCTTGAACAGGGTGGCCAAATC 240
Qy 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAAAGTCACTGAGCTATTGTGAACAGACT 300
Db 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAAAGTCACTGAGCTATTGTGAACAGACT 300
Qy 301 CGACGGTCATGACAGCTGTGACAGATTTTAATAATCTGCCGGACAGATTTAAAGACTTTC 360
Db 301 CGACGGTCATGACAGCTGTGACAGATTTTAATAATCTGCCGGACAGATTTAAAGACTTTC 360
Qy 361 TCTTGTTATTGAGATGCCGGAATTTACTCGCTGCTTATAGATCAACCGAAGAAATGTGCAA 420
Db 361 TCTTGTTATTGAGATGCCGGAATTTACTCGCTGCTTATAGATCAACCGAAGAAATGTGCAA 420
Qy 421 AGAAGCCCTTCTTACTATTGCGGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 480
Db 421 AGAAGCCCTTCTTACTATTGCGGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 480
Qy 481 CAATTCCGGAGTCTTGGGCGGAGAAACCAAGTAGGGAACAGACAGTAGTGAAGGTCT 540
Db 481 CAATTCCGGAGTCTTGGGCGGAGAAACCAAGTAGGGAACAGACAGTAGTGAAGGTCT 540
Qy 541 TCTTGTTGGGCAAGACACCCAGAGGACACCCCTGACTTTCGGACATGCTTAAGT 600
Db 541 TCTTGTTGGGCAAGACACCCAGAGGACACCCCTGACTTTCGGACATGCTTAAGT 600
Qy 601 TTGAGAGTGACAAAGCACCAGGACATCTCTCATGTGGAATATAGAGACACATTTCTCAACC 660
Db 601 TTGAGAGTGACAAAGCACCAGGACATCTCTCATGTGGAATATAGAGACACATTTCTCAACC 660
```



```
QY 661 TGTCCCTGAAGGAGTCTGTTCTTCTTAGTGGGTGAGCACTTCTCTGTCCAGAGCGCAGAGT 720
Db 661 TGTCCCTGAAGGAGTCTGTTCTTCTTAGTGGGTGAGCACTTCTCTGTCCAGAGCGCAGAGT 720
QY 721 TTGTCTTCAAGGGGATGATGACGTGTTTGTGAACCCCATCATCTTAACTTGA 780
Db 721 TTGTCTTCAAGGGGATGATGACGTGTTTGTGAACCCCATCATCTTAACTTGA 780
QY 781 ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGTGTGATGCCAATGCTG 840
Db 781 ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGTGTGATGCCAATGCTG 840
QY 841 GGCTCACCAGGTAAGAACTGAAGTACTACATCCAGAACTCTCTACACCGGCTCT 900
Db 841 GGCTCACCAGGTAAGAACTGAAGTACTACATCCAGAACTCTCTACACCGGCTCT 900
QY 901 ACCCAGCGTATGCCGGGGTGGTGGATTCTGTACTCCGCCCCCTTGCCTTGAGGCTGT 960
Db 901 ACCCAGCGTATGCCGGGGTGGTGGATTCTGTACTCCGCCCCCTTGCCTTGAGGCTGT 960
QY 961 ACAGTGGGACTAGCCGGTCCATCTCTACCCATATGATGATGTTTATACGGGAATGTGC 1020
Db 961 ACAGTGGGACTAGCCGGTCCATCTCTACCCATATGATGATGTTTATACGGGAATGTGC 1020
QY 1021 TTCAGAACTGGGCTTGTTCAGAGAGCAGCAAGGCTTCAGGACATTTGATTTGAAG 1080
Db 1021 TTCAGAACTGGGCTTGTTCAGAGAGCAGCAAGGCTTCAGGACATTTGATTTGAAG 1080
QY 1081 AGAAAAATAAGAAAAATATTTGTTCTTATAGACCTTAATGTTAGTACATAGCAGAAAA 1140
Db 1081 AGAAAAATAAGAAAAATATTTGTTCTTATAGACCTTAATGTTAGTACATAGCAGAAAA 1140
QY 1141 CTCAGAGATGATGATATCTGGTCTCAGTTCAGTGCAGGCTCCTAATTTAAATGCTGA 1196
Db 1141 CTCAGAGATGATGATATCTGGTCTCAGTTCAGTGCAGGCTCCTAATTTAAATGCTGA 1196

RESULT 3
US-10-109-563-1
; Sequence 1, Application US/10109563
; Publication No. US20020170079A1
; GENERAL INFORMATION:
; APPLICANT: Phillips, Michael W.
; APPLICANT: Leviten, Russell
; TITLE OF INVENTION: TRANSGENIC MICE COMPRISING
; TITLE OF INVENTION: BETA-1.3-N-ACETYL LACTOSAMINE DISRUPTIONS
; FILE REFERENCE: R-021
; CURRENT APPLICATION NUMBER: US/10/109,563
; CURRENT FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: US 60/280,706
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1260
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-109-563-1

Query Match 99.7%; Score 1192.8; DB 13; Length 1260;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1194; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGGCTGGAAGAGTCAGTGTGCTGGGCATCCTCATGATGSCAATGCTCT 60
Db 28 AGATGAGTGTGGGGCTGGAAGAGTCAGTGTGCTGGGCATCCTCATGATGSCAATGCTCT 87
QY 61 TCATTTATTTGATTTGGAAGTCTCCAAAAACAGTAGCCAGACAAAAATGGAAGGGAG 120
Db 88 TCATTTATTTGATTTGGAAGTCTCCAAAAACAGTAGCCAGACAAAAATGGAAGGGAG 147
QY 121 GAGTAATAATCCGAAAGAGAGTCTTGGAGGCCACCCAGCACTCCCGGGGCATCTGGA 180
Db 121 GAGTAATAATCCGAAAGAGAGTCTTGGAGGCCACCCAGCACTCCCGGGGCATCTGGA 180
```

RESULT 4
US-09-804-357-3

```
Db 148 GAGTAATAATCCGAAAGAGAGTCTTGGAGGCCACCCAGCACTCCCGGGGCATCTGGA 207
QY 181 ACAGGGAACAGAGAGAGCTGAACAGGTGGTACAATCCCATCTTGAACAGGGTGGCCAAATC 240
Db 208 ACAGGGAACAGAGAGAGCTGAACAGGTGGTACAATCCCATCTTGAACAGGGTGGCCAAATC 267
QY 241 AGACAGGGAGGTAGCCACATCTCCAAACACAAGTCACTGAGCTATTTGTGAACAGACT 300
Db 268 AGACAGGGAGGTAGCCACATCTCCAAACACAAGTCACTGAGCTATTTGTGAACAGACT 327
QY 301 CCACGGTCAIACACCTGTGACAGATTTTAAATATCTGCCGACAGATTTAAAGACTTTC 360
Db 328 CGACGGTCAIACACCTGTGACAGATTTTAAATATCTGCCGACAGATTTAAAGACTTTC 387
QY 361 TCTTGTATTGAGATGCCGGAATTACTCGCTGTATAGATCAACCCGAAAGAAATGTGCAA 420
Db 388 TCTTGTATTGAGATGCCGGAATTACTCGCTGTATAGATCAACCCGAAAGAAATGTGCAA 447
QY 421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTGCCAGAGGCAAG 480
Db 448 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTGCCAGAGGCAAG 507
QY 481 CAATTTCGGGAGTCTTGGGGCCGAGAAACCAACGTAGGGAACACAGACTAGTGGGTCT 540
Db 508 CAATTTCGGGAGTCTTGGGGCCGAGAAACCAACGTAGGGAACACAGACTAGTGGGTCT 567
QY 541 TCTGTTTGGGCAAGACACCCCCAGAGGACACCCCTGACCTTTCGGACATGCTTAAGT 600
Db 568 TCTGTTTGGGCAAGACACCCCCAGAGGACACCCCTGACCTTTCGGACATGCTTAAGT 627
QY 601 TTGAGAGTGACAAAGCACCAGGACATCCTCATGTGGAACATATAGAGACACATTTCTCAAC 660
Db 628 TTGAGAGTGACAAAGCACCAGGACATCCTCATGTGGAACATATAGAGACACATTTCTCAAC 687
QY 661 TGTCCCTGAAGGAGTGTGTTTCTTAGTGGGTGAGCACTTCTGTCCAGAGCGAGAGT 720
Db 688 TGTCCCTGAAGGAGTGTGTTTCTTAGTGGGTGAGCACTTCTGTCCAGAGCGCAGAGT 747
QY 721 TTGTCTTCAAGGGCGATGATGACGTGTTTGTGAACCCCATCATCTTAACTTGA 780
Db 748 TTGTCTTCAAGGGCGATGATGACGTGTTTGTGAACCCCATCATCTTAACTTGA 807
QY 781 ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGTGCAGCTGATCCACATGCTG 840
Db 808 ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGTGCAGCTGATCCACATGCTG 867
QY 841 GGCTCACCAGGATGAAGAACTGAAGTACTACATCCAGAGAGTCTTTACACCGGCTCT 900
Db 868 GGCTCACCAGGATGAAGAACTGAAGTACTACATCCAGAGAGTCTTTACACCGGCTCT 927
QY 901 ACCACCGGTATGCCGGGGTGGTGGATTCTCTACTCCGGCCCCCTTGCCTTGAGGCTGT 960
Db 928 ACCACCGGTATGCCGGGGTGGTGGATTCTCTACTCCGGCCCCCTTGCCTTGAGGCTGT 987
QY 961 ACAGTGCAGTACGCCGGGTCCATCTCTACCCATTTGATGATGTTTATACGGGAATGTGCC 1020
Db 988 ACAGTGCAGTACGCCGGGTCCATCTCTACCCATTTGATGATGTTTATACGGGAATGTGCC 1047
QY 1021 TTCAGAACTGGGCTTGTTCAGAGAGCAAGGCTTCAGGACATTTGATTTGAAG 1080
Db 1048 TTCAGAACTGGGCTTGTTCAGAGAGCAAGGCTTCAGGACATTTGATTTGAAG 1107
QY 1081 AGAAAAATAAGAAAAATATTTGTTCTTATAGACCTTAATGTTAGTACATAGCAGAAAA 1140
Db 1108 AGAAAAATAAGAAAAATATTTGTTCTTATAGACCTTAATGTTAGTACATAGCAGAAAA 1167
QY 1141 CTCAGAGATGATGATGATCTGCTCAGTTCGAAAGTCTTAAATTTAAATGCTGA 1196
Db 1168 CTCAGAGATGATGATGATCTGCTCAGTTCGAAAGTCTTAAATTTAAATGCTGA 1223
```



```
QY 123 GTAATAATCCCGAAGAGAGTCTTGGAGGACACCCAGCACTCCCGGGCATACTGGAC 182
Db 121 GTAATAATCCCGAAGAGAGTCTTGGAGGACACCCAGCACTCCCGGGCATACTGGAC 180
QY 183 AGGACAGGAGAGTCTGACAGGTGGTACATCCATCTTGAACAGGGTGGCCAACTAG 242
Db 181 AGGACAGGAGAGTCTGACAGGTGGTACATCCATCTTGAACAGGGTGGCCAACTAG 240
QY 243 ACAGGGAGCTAGCCACATCTCCAAACACAGTCACTGAGTATGTGAACAGCACTCG 302
Db 241 ACAGGGAGCTAGCCACATCTCCAAACACAGTCACTGAGTATGTGAACAGCACTCG 300
QY 303 ACGGTATGACAGCTGTGACAGATTTTATATCTGCGGAGACAGATTTAAAGACTTTC 362
Db 301 ACGGTATGACAGCTGTGACAGATTTTATATCTGCGGAGACAGATTTAAAGACTTTC 360
QY 363 TTCTATTGAGATGCGGGAATTACTCCCTGCTTATAGATCAACCGAAGAAATGTCAAG 422
Db 361 TTCTATTGAGATGCGGGAATTACTCCCTGCTTATAGATCAACCGAAGAAATGTCAAG 420
QY 423 AAGCCCTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAGCA 482
Db 421 AAGCCCTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAGCA 480
QY 483 ATTCGGAGCTTGGGCGGAGAAACCAAGTAGGAGACACAGTAGTAGGCTCTTC 542
Db 481 ATTCGGAGCTTGGGCGGAGAAACCAAGTAGGAGACACAGTAGTAGGCTCTTC 540
QY 543 CTGTTGGGCAAGACACCCCGAGGACACACCCCTGACCTTTCGGACATGCTTAAGTT 602
Db 541 CTGTTGGGCAAGACACCCCGAGGACACACCCCTGACCTTTCGGACATGCTTAAGTT 600
QY 603 GAGAGTGACAGCAGGACATCTCATGTGGAATATAGACACATTTCTCAACCTG 662
Db 601 GAGAGTGACAGCAGGACATCTCATGTGGAATATAGACACATTTCTCAACCTG 660
QY 663 TCCTGAAGAGAGTGTCTTCTAGTGGTGAGCACTTCTGTCACACACAGATTT 722
Db 661 TCCTGAAGAGAGTGTCTTCTAGTGGTGAGCACTTCTGTCACACACAGATTT 720
QY 723 GTCTTCAAGGCGATGATGACGTGTTTGTGAACACCCATCACATCCTTAATTTGAAT 782
Db 721 GTCTTCAAGGCGATGATGACGTGTTTGTGAACACCCATCACATCCTTAATTTGAAT 780
QY 783 AGCTTATCAAGAGCAAGCAAGACTTGTTCATAGTGAGTGATGCACATGCTGG 842
Db 781 AGCTTATCAAGAGCAAGCAAGACTTGTTCATAGTGAGTGATGCACATGCTGG 840
QY 843 CCTACCGGATAGAAACTGAAGTACTACATCCAGAGTCTTCTACACCGGCTCTAC 902
Db 841 CCTACCGGATAGAAACTGAAGTACTACATCCAGAGTCTTCTACACCGGCTCTAC 900
QY 903 CCACCGTATCGCGGGGTGGTGATTCCTGTACTCGGGCCCCCTTGCCTTGAAGGTGAC 962
Db 901 CCACCGTATCGCGGGGTGGTGATTCCTGTACTCGGGCCCCCTTGCCTTGAAGGTGAC 960
QY 963 AGTGCAGACTAGCGGGTCCATCTACCTATGTAGTATGATGTTTATACGGGAATGCGCTT 1022
Db 961 AGTGCAGACTAGCGGGTCCATCTACCTATGTAGTATGATGTTTATACGGGAATGCGCTT 1020
QY 1023 CAGAAACTGGGCTTCTCCAGAGAGCACAAAGGCTTCAGGACATTTTGTATTTGAAG 1082
Db 1021 CAGAAACTGGGCTTCTCCAGAGAGCACAAAGGCTTCAGGACATTTTGTATTTGAAG 1080
QY 1083 AAAAATAAGAAAAATATTTGTTCTATATAGACTAAAGTAGTACATAGCAGAAAACT 1142
Db 1081 AAAAATAAGAAAAATATTTGTTCTATATAGACTAAAGTAGTACATAGCAGAAAACT 1140
QY 1143 CAAGAGATGATGATATCTGGTCTCAGTTGCAAAAGTCTCAATTTAAATATGC 1193
Db 1141 CAAGAGATGATGATATCTGGTCTCAGTTGCAAAAGTCTCAATTTAAATATGC 1191
```

```
RESULT 6
US-09-804-357-13
; Sequence 13, Application US/09804357
; Patent No. US2001002480A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 1707
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (246)...(1436)
; NAME/KEY: misc.feature
; LOCATION: (1)...(1707)
; OTHER INFORMATION: n = A,T,C or G
US-09-804-357-13
```

```
Query Match 75.8%; Score 906.4; DB 9; Length 1707;
Best Local Similarity 84.9%; Pred. No. 4.3e-275;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGCGTGAAGAGTCAAGTTGCTGGGCATCCTGATGATGCAATGTCT 60
Db 244 AAATGAGTGTGGGCGTGAAGAAATAAGTTGTTGGGTATCTCTGATGATGCAATGTCT 303
QY 61 TCATTATTGATGTGGAAGTCTCCAAACACAGTAGCAAGCAAAATGGAAGGAG 120
Db 304 TCATTATTATTATGGAAGTCTCCAAACACAGTAGCAAGCAAAATGGAAGGAG 363
QY 121 GAGTAATAATCCGAAAGAGAAAGTTCTGGAAGCCACCCAGCACTCCCGGGCATACTGGA 180
Db 364 AAGTAATAATCCGAAAGAGAAAGTTCTGGAAGATATCTACCCCTCCCGAGGCATCTGGA 423
QY 181 ACAGGGAACAGAGAGAGTGAACAGGTGCTACATCCCATCTTGAACAGGTGGCCATC 240
Db 424 ACCGAGACAAGAGAGAGTGAACCCGGCAGTGAACCCCATCTCTGAGCATGCTGACCAAC 483
QY 241 AGACAGGGAGTAGCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACT 300
Db 484 AGACAGGGAGGGCGGAGGCTCTCCATATAGCCATCTGCACTACTGCAACCTGACC 543
QY 301 CGACGGTCATGACAGCTGTGACAGATTTTAAATATCTGCCGAGACAGATTTAAAGACTTC 360
Db 544 TGAGGGTCACGTCGGTGGTTACGGGTTTTAAACAACTTCCCGGACAGATTTAAAGACTTC 603
QY 361 TCTTGTATTGAGATGCGGGAATTACTGCTGCTTATAGATCAACCGAAGAAATGTGCA 420
Db 604 TGCTGTATTGAGATGCGGCAATTTATTCCTGCTATAGATCAAGCCGATTAAGTGTGCA 663
QY 421 AGAAGCCCTCTTACTATTGGCGATAAAGTCCCTCATTTCCACATTTTGGCAAGAGCAAG 480
Db 664 AGAAGCCCTTCTTGTGCTGGGATTAAGTCCCTCACTCCACATTTTGGCAAGAGCAAG 723
QY 481 CAATTCGGGAGTCTTGGGGCCGAGAAACCAACGCTAGGGAACACAGACAGTACTAGGAGTCT 540
Db 724 CAATCCGGGAATCTTGGGGCCGAGAAAGCAACGCGAGGGAACCAACAGGTGTGTGAGTCT 783
```

```
QY 541 TCCTGTTGGCAGACACACCCAGAGGACCAACCCCTGACCTTTTCGGACATGCTTAAGT 600
DB 784 TCCTGCTGGCCAGACACCCAGAGGACCAACCCCGACCTTTTCAGATATGCTGAAT 843
QY 601 TTGAGAGTGCACAGCACCAGGACATCCTCATGTGGAAGTATAGACACACATTTTCACCC 650
DB 844 TTGAGAGTGCAGAGCACCAGACATCTTATGTGAAGTACAGACACATTTTCACCACT 903
QY 661 TCTCCCTGAAGAGTGCCTGTTTCATAGTGGTGAGCAGCTTCCTGACAGCAGAGT 720
DB 904 TGTCTCTGAAGAGTGCCTGTTTCAGGTGGTGAAGTACATTCCTGCCACACACTGAGT 963
QY 721 TTGCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCACATCCTTAATTACTTGA 780
DB 964 TTGTTTCAAGGGCGATGATGATGTTTGTGAACACCCATCACATCCTGAATTACTTGA 1023
QY 781 ATAGCTTTATCCAGAGCAAGCAAGACTTCTTATAGGTGACGTGATCCACAAATGCTG 840
DB 1024 ATAGTTTATCCAGAGCAAGCAAGACTTCTTATAGGTGATGATCCACAAATGCTG 1083
QY 841 GGCTTCACCGGGATGAAGAACTGAAGTACTACATCCAGAGTCTTCTACACCGGCTCT 900
DB 1084 GACCTCATCGGATGAAGAGCTGAAGTACTACATCCAGAGTCTTCTACCTGCGCTCT 1143
QY 901 ACCACCGTATCCCGGGGTGTGGATTCCTGTACTCCGGCCCTTGCCTTGAGGCTGT 960
DB 1144 ACCACCGTATCCCGGGGTGTGGATTCCTGTACTCCGGCCCTTGCCTTGAGGCTGT 1203
QY 961 ACAGTCCGACTAGCCGGGTCCATCTCTACCTATTGATGATGTTTATACGGGAATGTGCC 1020
DB 1204 ACATATCACTGACAGGTCATCTCTACCCATTGATGACGTTTATGGAATGTGCC 1263
QY 1021 TFCAGAACTGGCCCTTCTCCAGAGACCAAAAGGCTTCAGGACATTTGATATTGAAG 1080
DB 1264 TFCAGAACTGGCCCTTCTCCAGAGACCAAAAGGCTTCAGGACATTTGATATTGAAG 1323
QY 1081 AGAAAAATGAAGAAATATTTGCTTATATAGACCTATGTTAGTACATAGCAAGAAAC 1140
DB 1324 AGAAAAATGAAGAAATATTTGCTTATATAGACCTATGTTAGTACATAGCAAGAAAC 1383
QY 1141 CTCAGAGATGATGATATCTGCTCTCAGTTCGAAGTCTTAATTTAAATGCTGA 1196
DB 1384 CTCAGAGATGATGATATCTGCTCTCAGTTCGAAGTCTTAATTTAAATGCTGA 1439
```

RESULT 7

```
US-09-804-006-13
; Sequence 13, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/126001
; CURRENT APPLICATION NUMBER: US/09/804, 006
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/292,228
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 1707
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (246)...(1436)
; NAME/KEY: misc_feature
```

```
; LOCATION: (1)...(1707)
; OTHER INFORMATION: n - A,T,C or G
US-09-804-006-13

Query Match
Best Local Similarity 84.9%; Score 906.4; DB 10; Length 1707;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 1 AGATGAGTGGGGCGTGAAGAGTCAAGTTGCTGGGCATCCGTGATGATGGCAAAATGTCT 60
DB 244 AAATGAGTGTGGACGTCGAAGAATAAAGTTGTTGGTATCTCTCATGATGCCAAATGTCT 303
QY 61 TCATTTATTTGATTTGCGAAAGTCTCCAAAACAGTAGCCAAAGCAAAAATGGAAGGGAG 120
DB 304 TCATTTATTTATTTATGGAAGTCTCCAAAAGAGTAGCCAAAGCAAAAATGGAAGGGG 363
QY 121 GAGTAATAATCCCGAAAGAGAAGTTCTGGAAGCCACCCAGCACTCCCGGGCATACTGGA 180
DB 364 AAGTAATAATACCCAAAGAGAAGTTCTGGAAGATATCTACCCCTCCCGAGGCATACTGGA 423
QY 181 ACAGGGAACAGGAGAAGCTGAACAGTGTGACATCCCATCTTGAACAGGTTGCCAAATC 240
DB 424 ACCAGAGCAAGAGAGTGAACCGGAGTACAAACCCATCTCTGAGCTGCTGACCAACC 483
QY 241 AGACAGGGAGCTAGCCACATCTCCAAACACAAAGTCACTACCTGAGCTATTGTGAACCACT 300
DB 484 AGACGGGGAGGGGGCGGCTCTCCAAATATAAGCCATCTGAACACTCTGGAACCTGACC 543
QY 301 CGAGGTCATGACAGCTGTGACAGATTTTAAATATCTGCCGGACAGATTTTAAAGACTTTC 360
DB 544 TGAGGTCAGCTCGGTGGTTACCGGGTTTAAACACTTGGCCGACAGATTTTAAAGACTTTC 603
QY 361 TCTTGTATTTGAGATGCCGAATTTACTCGCTGTTATAGATCAACCGCAAAATGTGCAA 420
DB 604 TGTGTATTTGAGATGCCGAATTTACTCTGCTTATAGATCAACCGGATGATGTGCAA 663
QY 421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTTCCACATTTTGCAGAGGCAAG 480
DB 664 AGAAGCCCTTCTTCTGCTGCGATTAAGTCCCTCACTCCACATTTTGCAGAGGCAAG 723
QY 481 CAATTCGGAGTCTTGGGGCCGAGAACCAACAGTGTAGGGAACCAACAGAGTAGTACAGGCTCT 540
DB 724 CAATCCGGGAATCTTGGGGCCAGAAAGCAACGCGGGAACCAACAGGTCGCGAGTCT 783
QY 541 TCCTGTTGGGCAAGACACCCCGAGGACCAACCCCTGACCTTTCGGACATGCTTAAAT 600
DB 784 TCCTGCTGGCCAGACACCCCGAGGACCAACCCCGACCTTTCAGATATGCTGAAT 843
QY 601 TTGAGAGTGCACAGCACCAGGACATCCTCATGTGGAAGTATAGACACACATTTCTCAACC 660
DB 844 TTGAGAGTGCAGAGCACCAGCAATCTTATGGAAGTACAGAGACACTTTCTTCAACT 903
QY 561 TGTCCCTGAAGGAAGTGTCTTCTTAGTGGGTGAGCACTTCTCTGACAGAGCAGAGT 720
DB 904 TGTCTCTGAAGGAAGTGTCTTCTCAGGTGGGTAACTACTTCTGCCAGACACTGAGT 963
QY 721 TTGCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCAATCCTTAATTACTTGA 780
DB 964 TTGTTTCAAGGGCGATGAGCATGTTTGTGAACACCCATCAATCCTGAATTACTTGA 1023
QY 781 ATAGCTTATCCAGAGCAAGCAAGACTTGTTCATAGGTGAGCTGATCCAAATGCTG 840
DB 1024 ATAGTTTATCCAGAGCAAGCAAGACTTCTTATAGGTGATGATCCAAATGCTG 1083
QY 841 GGCTTCACCGGGATGAAGAACTGAAGTACTACATCCAGAGTCTTCTACACCGGCTCT 900
DB 1084 GACCTCATCGGATGAAGAGTGAAGTACTACATCCAGAGTGTGTTACTCTGGGCTCT 1143
QY 901 ACCACCGTATCCCGGGGTGTGGATTTCTCTGTTACTCCGGCCCTTGCCTTGAGGCTGT 960
DB 1144 ACCACCGTATCCCGGGGTGTGGATTTCTCTACTCCGGCCCTTGCCTTGAGGCTGT 1203
QY 961 ACAGTCCGACTAGCCGGGTCCATCTCTACCTATTGATGATGTTTATACGGGAATGTGCC 1020
```

```
Db 1204 ACCATATCACTGACAGGTCCTCTCTACCCCATGATGACGCTTTATATCTGGAATGTGCC 1263
QY 1021 TTCAAGAACTGGGCTGTGTCCAGAGAACCAAAAGGCTTCAGGACATTTGATTTGAAG 1080
Db 1264 TTCAAGAACTGGGCTGTGTCCAGAGAACCAAAAGGCTTCAGGACATTTGATTCGAGG 1323
QY 1081 AGAAAATTAAGAAATATTTGTTCTATATAGACCTAATGTTAGTACATACGAAAC 1140
Db 1324 AGAAAACAAAATAACATCTCTCTATGTAGATCTGATGTAGTACATAGTAAAC 1383
QY 1141 CTCAGAGATGATTGATATCTGTCTCAGTTCGAAGTCTTAATTTAAATGCTGA 1196
Db 1384 CTCAGAGATGATTGATATTTGTTCTCAGTTCGAGAGTCTCATTTTAAATGCTAA 1439

RESULT 8
US-10-109-563-3
; Sequence 3, Application US/10109563
; Publication No. US20020170075A1
; GENERAL INFORMATION:
; APPLICANT: Leviten, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE COMPRISING
; TITLE OF INVENTION: BETA-1,3-N-ACETYLACTOSAMINE DISRUPTIONS
; FILE REFERENCE: R-021
; CURRENT APPLICATION NUMBER: US/10/109,563
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: US 60/280,706
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1831
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-109-563-3

Query Match 75.8%; Score 906.4; DB 13; Length 1831;
Best Local Similarity 84.9%; Pred. No. 4.5e-275;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGCTCCAGAGTCAAGTCTCTGGGCTCTCTGATGATGGCAATGTCT 60
Db 234 AATGAGTGTGGGCTCCAGAGTCAAGTCTCTGGGCTCTCTGATGATGGCAATGTCT 293
QY 61 TCATTATTGTTGTTGGAAGTCTCCAAAACAGTAGCCCAAGCAAAAATGGAAGGGAG 120
Db 294 TCATTATTATTATGGAAGTCTCCAAAACAGTAGCCCAAGCAAAAATGGAAGGGG 353
QY 121 GAGTAATAATCCGAAAGAGAGTCTTGGAAAGCCACCCAGCACTCCCGGGCATACTGGA 180
Db 354 AGTAATAATACCCAAAGAGAGTCTTGGAAAGATATCTACCCCTCCCGAGGCACTGGA 413
QY 181 ACAGGGAACAGAGAGTCTTGGAAAGTCTTGGAAAGTCTTGGAAAGTCTTGGAAAGT 240
Db 414 ACCGAGCAAGAGAGTCTTGGAAAGTCTTGGAAAGTCTTGGAAAGTCTTGGAAAGT 473
QY 241 AGACAGGGAGTCTTGGAAAGTCTTGGAAAGTCTTGGAAAGTCTTGGAAAGTCTTGG 300
Db 474 AGACAGGGAGTCTTGGAAAGTCTTGGAAAGTCTTGGAAAGTCTTGGAAAGTCTTGG 533
QY 301 CGAGGTCATGAGAGTCTTGGAAAGTCTTGGAAAGTCTTGGAAAGTCTTGGAAAGTCT 360
Db 534 TGAGGTCATGAGAGTCTTGGAAAGTCTTGGAAAGTCTTGGAAAGTCTTGGAAAGTCT 593
QY 361 TCTGTTATTTGAGATCCGGAATCTCGTCTGTTATAGATCAACGAGAAATGTGCA 420
Db 594 TGTGTTATTTGAGATCCGGAATCTCGTCTGTTATAGATCAACGAGAAATGTGCA 653
QY 421 AGAAGCCCTTCTTACTATTGGGCAATAAGTCCCTCATTCACATTTTGGCAGAGGCAAG 480
Db 654 AGAAGCCCTTCTTACTATTGGGCAATAAGTCCCTCATTCACATTTTGGCAGAGGCAAG 713
```

```
QY 481 CAATTCGGGAGTCTTGGGGCGAGAAACCAACAGTAGGGAACCAACAGACAGTCTGAGGGTCT 540
Db 714 CAATTCGGGAGTCTTGGGGCGAGAAACCAACAGTAGGGAACCAACAGTCTGAGGGTCT 773
QY 541 TCCTGTTGGGCAAGACACCCCGAGAGAACCAACAGTCTTGGGACATGCTTAAAGI 600
Db 774 TCCTGTTGGGCAAGACACCCCGAGAGAACCAACAGTCTTGGGACATGCTTAAAT 833
QY 601 TTGAGAGTGACAAGCACCCCGAGAGAACCAACAGTCTTGGGACATGCTTAAAC 660
Db 834 TTGAGAGTGACAAGCACCCCGAGAGAACCAACAGTCTTGGGACATGCTTAAAC 893
QY 661 TGTCCCTGAAGAGTCTGTTTCTTAGTGGTGGGAGCATTCTCTGCGACAGCAGAGT 720
Db 894 TGTCTCTGAAGAGTCTGTTTCTTAGTGGTGGGAGCATTCTCTGCGACAGCAGT 953
QY 721 TTGCTTCAAGGGGATGATGACGTTTGTGACACCCATCAGATCTCTTAACTTGA 780
Db 954 TTGCTTCAAGGGGATGATGACGTTTGTGACACCCATCAGATCTCTTAACTTGA 1013
QY 781 ATAGCTTATCCAAGAGCAAGCCCAAGAGTCTTGTATAGTGTGATGATGATGATGAT 840
Db 1014 ATAGCTTATCCAAGAGCAAGCCCAAGAGTCTTGTATAGTGTGATGATGATGATGAT 1073
QY 841 GGCCTCACCGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTGTATAGTGTGAT 900
Db 1074 GACCTCATCGGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTGTATAGTGTGAT 1133
QY 901 ACCCAGCTATGCGGGGGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 960
Db 1134 ACCCAGCTATGCGGGGGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1193
QY 961 ACAGTGCAGTACGCGGGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1020
Db 1194 ACCATATCACTGACAGGTCCTCTTACCCATTTGATGACGCTTTATATCTGGAATGTGCC 1253
QY 1021 TTCAGAACTGGGCTTGTTCAGAGAGCAACAAAGGCTTCAGAGACATTTGATATCGAGG 1080
Db 1254 TTCAGAACTGGGCTTGTTCAGAGAGCAACAAAGGCTTCAGAGACATTTGATATCGAGG 1313
QY 1081 AGAAAATTAAGAAATATTTGTTCTATATAGACCTAATGTTAGTACATAGCAGAAAC 1140
Db 1314 AGAAAATTAAGAAATATTTGTTCTATATAGACCTAATGTTAGTACATAGTACAGAAAC 1373
QY 1141 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1196
Db 1374 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1429
```

```
RESULT 9
US-09-972-912-1
; Sequence 1, Application US/09972912
; Patent No. US20020110867A1
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R.
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/972,912
```



```
;
;
;   REGISTRATION NUMBER: 36,688
;   REFERENCE/DOCKET NUMBER: 1488.0620001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 371-2600
;   TELEFAX: (202) 371-2540
;   INFORMATION FOR SEQ ID NO: 20:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 857 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: both
;   TOPOLOGY: both
;   MOLECULE TYPE: cDNA
;   SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-972-912-20

Query Match      24.1%; Score 288.8; DB 10; Length 857;
Best Local Similarity 85.0%; Pred. No. 3.6e-80;
Matches 345; Conservative 0; Mismatches 59; Indels 2; Gaps 2;

QY 787 TATCCAGAGCAAGCCAAAGACTGTTTCATAGTACGTGATCCACAATGCTGGGCGCTC 846
Db 20 TATTCAGAGCAAGCCAAAGACTGTTTCATAGTACGTGATCCACAATGCTGGGCGCTC 79
QY 847 ACCGGGATAGAACTGAAGTACTACATCCCAAGTCTTCTACACGGCGCTACCCCA 906
Db 80 ATCGGGATAGAAAGCTGAAGTACTACATCCCAAGTCTTCTACTCJGGCGCTACCCAC 139
QY 907 CGTATCGCGGGGTGGTGGATTCCTGTACTCGGGCCCTTGGCTTGAAGCTGTACAGTG 966
Db 140 CCTATCGAGGGAGGGGGGTCTCTACTCCGGCCACCTGGCGCTGAGGCTGTACCATA 199
QY 967 CGACTAGCGGGTCCATCTCTACCTCTATTGATGATCTTTATACGGGAATGTGCCCTCAGA 1026
Db 200 TCACTACACAGGTCATCTCTACCTCTATTGATGATCTTTATACGGGAATGTGCCCTCAGA 259
QY 1027 AACTGGGCGCTTGTCCAGAGCAACAAAGGCTTCAGGACATTTGATATTGAAGAGAAAA 1086
Db 260 AACTGGGCGCTTGTCCAGAGCAACAAAGGCTTCAGGACATTTGATATTGAAGAGAAAA 319
QY 1087 ATAGAGAAATATTGTTCTCTATAGACCTATGTTAGTACATACA-GAAGACCTCAA 1145
Db 320 AAAAAATACATCTCTCTCTATGATCTGATGTTAGACATAGNAGAGAAACCTCAA 379
QY 1146 GAGATGATTGATATCTGGTCTC-AGTTCAGAACTCTCAATTTAAA 1190
Db 380 GAGATGATTGATATTGGGCTCAAGNTGAGAGTGTCTCAATTTAAA 425

RESULT 11
US-09-972-912-26
; Sequence 26, Application US/09972912
; Patent No. US20020110867A1
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R.
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version: #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/972.912
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
```

```
;
;
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/049,022
;   FILING DATE: <Unknown>
;   ATTORNEY/AGENT INFORMATION:
;   NAME: STEFFE, ERIC K.
;   REGISTRATION NUMBER: 36,688
;   REFERENCE/DOCKET NUMBER: 1488.0620001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 371-2600
;   TELEFAX: (202) 371-2540
;   INFORMATION FOR SEQ ID NO: 26:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 480 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: both
;   TOPOLOGY: both
;   MOLECULE TYPE: cDNA
;   SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-09-972-912-26

Query Match      19.7%; Score 236; DB 10; Length 480;
Best Local Similarity 76.2%; Pred. No. 1.2e-63;
Matches 343; Conservative 0; Mismatches 97; Indels 10; Gaps 4;

QY 302 GACGGTCATGACAGCTGTGACAGATTTTAATTAATCTGCCGGACAGATTTAAGACTTTCT 361
Db 2 GAGGGTCAGCTCGGTGGTTACGGGTTTTAAACAACCTGCCGGACAGATTTAAGACTTTCT 61
QY 362 CTTGTATTGTAGATGCCGAATTTACTCGCTCTATAGATCAACCGAAGAAATGTGCAAA 421
Db 62 GCTGTATTGTAGATGCCGAATTTACTCGCTCTATAGATCAACCGGAGTAAGTGTGCAAA 121
QY 422 GAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTGCAGAGCAAGCAA-G 480
Db 122 GAAACCTTTTCTTGTGCTGGCGATTAAAGTCCCTCACTCCACATTTTGCAGAGCAAGG 181
QY 481 CAATTCCGGAGTCTTGGGGCCGAGAAACCAACCTAGGAGCAACAGACTAGTGGGTCT 540
Db 182 CAATCCGGAATCTTGGGGCCGAGAAACCAACCTAGGAGCAACCGGTGGTGGCGAGTCT 241
QY 541 TCCTGTGGGCAAGACACCCCGAGAGCAACCCCTGACCTTTCCGGACATGCT-TAAG 599
Db 242 TCCTGCTGGCGAGACACCCCGAGAGCAACCCCGACCTTTTCAGATATGCTGAAT 301
QY 600 TTTGAGATGCAAGCAGCAGGACATCTCATGTGG-----AACTATAGAGACACATCTT 655
Db 302 TTTGAGATGCAAGCAGCAGCAGCAGCATTTCTTGTGGGAACCTACAGAGGACACTTTCTTC 361
QY 656 CAACCTGTCCCTGAAGAGTGTCTGTTCTTAGTGGGTG-----AGCACTTCTCTGCCAG 711
Db 362 AANTTGTCTNTGGAGGAAGTGTCTGTTTTCAGGTGGGTTAAGTTATTTCTCTGCCAG 421
QY 712 ACGCAGAGTTTGTCTTCAAGGGCGGATGATG 741
Db 422 ACATTGAGTTTGTCTTTTCAAGGGCGGATG 451

RESULT 12
US-09-972-912-11
; Sequence 11, Application US/09972912
; Patent No. US20020110867A1
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R.
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
```



```
;
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/972,912
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/049,022
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0620001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-972-912-11

Query Match      18.8%; Score 224.6; DB 10; Length 282;
Best Local Similarity 86.9%; Pred. No. 3.3e-60;
Matches 245; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 464 TTTTCCAGAGGCAAGCAATTCGGAGTCTTGGGCGGAGAACCAACGTAAGGAACCA 523
Db 1 TTTTCCAGAGGCAAGCAATTCGGAGTCTTGGGCGGAGAACCAACGTAAGGAACCA 60

Qy 524 GACAGTAGTGAGGCTTCTCTGTTGGGCAAGACACCCCGAGAGACATCTCATGTGGAACCTATAG 583
Db 61 AACGGTGTGGAGTNTTCTGCTGGGCGGAGAACCAACGTAAGGAACCA 120

Qy 584 TTCCGACATCTTAAGTTTGAGAGTGACAAAGCAGACACCCCGAGAGACATCTCATGTGGAACCTATAG 643
Db 121 TTCAGATATCTGAAATTTGAGAGTGAGAAAGCAGACATCTTATGTGGAACCTACAG 180

Qy 644 AGACACATCTTCAACCTGCTCCCTGAGAGAGTGTCTTCTAGTGGTGGTGAGCACTTC 703
Db 181 AGACACATTTTCAACCTGCTCTGAGGAAGTGTCTTCTAGTGGTGGTGAGCACTTC 240

Qy 704 CTGTCCAGAGCAGAGTTTCTTCAAGGCGGATGATGACGT 745
Db 241 CTGCCCAGACACTGAGTTTCTTCAAGGCGGATGATGATGT 282

RESULT 13
US-09-972-912-34
; Sequence 34, Application US/09972912
; Patent No. US20020110867A1
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R.
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
;
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/972,912
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/049,022
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0620001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-972-912-34

Query Match      18.8%; Score 224.6; DB 10; Length 282;
Best Local Similarity 86.9%; Pred. No. 3.3e-60;
Matches 245; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 464 TTTTCCAGAGGCAAGCAATTCGGAGTCTTGGGCGGAGAACCAACGTAAGGAACCA 523
Db 1 TTTTCCAGAGGCAAGCAATTCGGAGTCTTGGGCGGAGAACCAACGTAAGGAACCA 60

Qy 524 GACAGTAGTGAGGCTTCTCTGTTGGGCAAGACACCCCGAGAGACATCTCATGTGGAACCTATAG 583
Db 61 AACGGTGTGGAGTNTTCTGCTGGGCGGAGAACCAACGTAAGGAACCA 120

Qy 584 TTCCGACATCTTAAGTTTGAGAGTGACAAAGCAGACACCCCGAGAGACATCTCATGTGGAACCTATAG 643
Db 121 TTCAGATATCTGAAATTTGAGAGTGAGAAAGCAGACATCTTATGTGGAACCTACAG 180

Qy 644 AGACACATCTTCAACCTGCTCCCTGAGAGAGTGTCTTCTAGTGGTGGTGAGCACTTC 703
Db 181 AGACACATTTTCAACCTGCTCTGAGGAAGTGTCTTCTAGTGGTGGTGAGCACTTC 240

Qy 704 CTGTCCAGAGCAGAGTTTCTTCAAGGCGGATGATGACGT 745
Db 241 CTGCCCAGACACTGAGTTTCTTCAAGGCGGATGATGATGT 282

RESULT 14
US-09-972-912-31
; Sequence 31, Application US/09972912
; Patent No. US20020110867A1
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R.
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```


APPLICATION NUMBER: US/09/972,912
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/049,022
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0620001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-972-912-31
Query Match 18.7%; Score 223.6; DB 10; Length 303;
Best Local Similarity 86.4%; Pred. No. 7.2e-60;
Matches 247; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 342 GACAGATTTAAAGACITTCCTCTGTTATTCAGATGCGGAAATCTCGCTGTATAGAT 401
Db 1 GACAGATTTAAAGACITTCCTCTGTTATTCAGATGCGGAAATCTCGCTGTATAGAT 60
QY 402 CAACCGAAGAAATGTGCAAGAAGCCCTCTTACTATGGCGATAAAGTCCCTCATTC 461
Db 61 CAGCCGATGAAGTGTGCAAGAAGCCCTCTTACTATGGCGATAAAGTCCCTCATTC 120
QY 462 CATTTGCCAAGGCAAGCAATTCGGAGTCTTGGGGCCGAGAAACCAACGTTAGGGAAC 521
Db 121 CATTTGCCAAGGCAAGCAATTCGGAGTCTTGGGGCCGAGAAACCAACGTTAGGGAAC 180
QY 522 CAGACAGTAGTGGGTCTTCTGTTGGGCAAGACACCCCGAGGACACACCCCTGAC 581
Db 181 CAACGGTGTGGGAGTCTTCTGTTGGGCAAGACACCCCGAGGACACACCCCGAC 240
QY 582 CTTTCGACATGCTTAAAGTTGAGGTGACAAAGCAGCAGGACATCC 627
Db 241 CTTTCAGATATGCTGAATTTGAGAGTTAGAGCACCAGACATTC 286

RESULT 15
US-09-972-912-12
Sequence 12, Application US/0972912
Patent No. US20020110867A1
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R.
RUBEN, STEVEN M.
TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.D.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/972,912
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/049,022
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0620001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 266 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-972-912-12
Query Match 17.7%; Score 211.2; DB 10; Length 266;
Best Local Similarity 86.9%; Pred. No. 5.5e-56;
Matches 231; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 464 TTTTGGCCAGAGCAAGCAATTCGGAGTCTTGGGGCCGAGAAACCAACGTTAGGGAACCA 523
Db 1 TTTTGGCCAGAGCAAGCAATTCGGAGTCTTGGGGCCGAGAAACCAACGTTAGGGAACCA 60
QY 524 GACAGTAGTGAGGTCTTCTGTTGGGCAAGACACCCCGAGGACACCCCTGACCT 583
Db 61 AACGGTGTGGGAGTNTCTCTGTTGGGCAAGACACCCCGAGGACACCCCTGACCT 120
QY 584 TTCGACATGCTTAAGTTTGAGAGTGACAAAGCAGGACATCTCTATGTGGAACATAG 643
Db 121 TTCAGATATGCTGAATTTGAGAGTNAGAAGCACCAGACATTTATGTGGAACATAG 180
QY 644 AGACATCTTCAACCTGTCCTGAAGGAAGTGTCTTAGTGGTGGTGGAGCACTTC 703
Db 181 AGACATCTTCAACCTGTCCTGAAGGAAGTGTCTTAGTGGTGGTGGAGCACTTC 240
QY 704 CTGTCCAGACGCGAGAGTTTGTCTTCA 729
Db 241 CTGTCCAGACGCGAGAGTTTGTCTTCA 266
Search completed: October 20, 2003, 22:44:22
Job time : 334.002 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 13:13:22 ; Search time 6381.7 Seconds
(without alignments)
10942.656 Million cell updates/sec

Title: US-09-804-357B-13
Perfect score: 1707
Sequence: 1 acgcgtccgcgcagcgcag.....aatctgttggatggccctt 1707

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_em:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rnd:*
- 36: em_htg_man:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1534.8	89.9	2811	9	BC047933	Homo sapi
2	1531.4	89.7	1831	9	AF092051	Homo sapi
3	1509.2	88.4	2714	9	AF288208	Homo sapi
4	1504.2	88.1	1912	6	BD093191	Useful po
5	1504.2	88.1	1912	9	AB049584	Homo sapi
6	1311.4	76.8	95088	9	AC093401	Homo sapi
7	1309.4	76.7	2454	9	BC030579	Homo sapi
8	1269.8	74.4	2516	9	AF288209	Homo sapi
9	1231.4	72.1	2442	6	BD156964	Primer fo
10	1231.4	72.1	2442	9	AK002009	Homo sapi
11	1181	69.2	1197	9	HSA6077	Homo sapi
12	1002.2	58.7	2672	10	AY043479	Mus muscu
13	954.2	55.9	200191	9	AL390027	Human DNA
14	954.2	55.9	288669	2	AL390027	Homo sapi
15	941.2	55.1	2468	10	BC009075	Mus muscu
16	940.2	55.1	210863	10	AL772364	Mouse DNA
17	931.4	54.6	1260	10	AF092050	Mus muscu
18	925.4	54.2	178009	2	AC119549	Rattus no
19	917.4	53.7	318862	2	AC109547	Rattus no
20	910.6	53.3	172509	2	AC024716	Homo sapi
21	717.2	42.0	832	6	BD149792	Primer fo
22	577.6	33.8	126480	2	AC040954	Mus muscu
23	533.6	31.3	556	6	BD071997	Secreted
24	425.2	24.9	170370	2	BX469915	Danio rer
25	423.2	24.8	2162	5	AF321831	Danio rer
26	414.2	22.3	197598	2	AC144887	Bos tauru
27	379	22.2	1682	5	AF321828	Danio rer
28	229.8	13.5	189326	9	AC018462	Homo sapi
29	223.4	13.1	473	6	BD052134	Sequence
30	203.4	11.9	1245	6	AX513603	Sequence
31	192.6	11.3	1564	5	AF321829	Danio rer
32	176.8	10.4	230799	2	AC097668	Rattus no
33	176.4	10.3	2260	5	AF321827	Danio rer
34	174.4	10.2	1134	6	AR202302	Sequence
35	174.2	10.2	209192	2	AC102609	Mus muscu
36	171	10.0	1502	10	AF502429	Mus muscu
37	171	10.0	2367	10	BC031187	Mus muscu
38	167.8	9.8	1360	6	AX451423	Sequence
39	166.2	9.7	1434	9	AF502430	Homo sapi
40	166.2	9.7	168880	9	AC017104	Homo sapi
41	164.8	9.7	2003	6	AX513601	Sequence
42	163.2	9.6	1116	6	BD063610	Human pro
43	163.2	9.6	1119	9	HSA130847	Homo sapi
44	163.2	9.6	1144	9	AF293973	Homo sapi
45	163.2	9.6	2186	6	BD063635	Human pro

ALIGNMENTS

RESULT 1
BC047933
LOCUS
DEFINITION
BC047933 Homo sapiens UDP-glucNAc:betaGal
beta-1,3-N-acetylglucosaminyltransferase 1, transcript variant 1,
mRNA (cDNA clone MGC:49838 IMAGE:5767816), complete cds.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
BC047933
MGC.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2811)
AUTHORS
Straussberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

183 CCCCGCCCGCCGAGCTGCTCCCGGACAGATATGAGAAATGAGTGTTGACG 242
184
260 TCGAAGAAATAAAGTTGTTGGTATCTCTGATGATGGCAATGCTCTCATTTATTTAT 319
261
243 TCGAAGAAATAAAGTTGTTGGTATCTCTGATGATGGCAATGCTCTCATTTATTTAT 302
244
320 GGAAGTCTCCAAAGCAGTAGCAAGAAAGAAATGGAAGAGGGAAGTATATACCCAA 379
321
303 GGAAGTCTCCAAAGCAGTAGCAAGAAAGAAATGGAAGAGGGAAGTATATACCCAA 362
304
380 AGAAGTCTCGAAGATATCTACCCCTCCCGAGGATAGTGAACCGGAGCAAGAGAA 439
381
363 AGAAGTCTCGAAGATATCTACCCCTCCCGAGGATAGTGAACCGGAGCAAGAGAA 422
364
440 GCTGAACCGGAGTACAAACCCATCTGAGCATGCTGACCAACGAGCGGGAGCGGG 499
441
423 GCTGAACCGGAGTACAAACCCATCTGAGCATGCTGACCAACGAGCGGGAGCGGG 482
424
500 CAGGCTCTCCAATATAAGCCATCTGAACCTACTCGGAACCTGACCTGAGGTCACGTCGGT 559
501
483 CAGGCTCTCCAATATAAGCCATCTGAACCTACTCGGAACCTGACCTGAGGTCACGTCGGT 542
484
560 GGTACCGGGTTTACAACTTGGCGACAGATTAAAGACTTTCTGCTGTTATGAGATG 619
561
543 GGTACCGGGTTTAAACACTTGGCGGACAGATTAAAGACTTTCTGCTGTTATGAGATG 602
544
620 CCGCAATTATTACCTGCTTATAGATCAGCCGGATAGTGTGCAAGAAAGCTTTCTTGTT 679
621
603 CCGCAATTATTACCTGCTTATAGATCAGCCGGATAGTGTGCAAGAAAGCTTTCTTGTT 662
604
680 GCTGGGATTAAGTCCCTCACTCCACATTTTGGCAGAGCAAGCAATCGGGAATCTTG 739
681
663 GCTGGGATTAAGTCCCTCACTCCACATTTTGGCAGAGCAAGCAATCGGGAATCTTG 722
664
740 GGGCAAGAAAGCAACCGAGGAAACCAAGCGTGGTCCGAGTCTTCTGCTGGCCGAC 799
741
723 GGGCAAGAAAGCAACCGAGGAAACCAAGCGTGGTCCGAGTCTTCTGCTGGCCGAC 782
724
800 ACCCCAGAGAGCAACCAACCCGACCTTTCAGATATGCTGAAATTTGAGAGTGAAGCA 859
801
783 ACCCCAGAGAGCAACCAACCCGACCTTTCAGATATGCTGAAATTTGAGAGTGAAGCA 842
784
860 CCAAGACATTTTAAGTGAACCTACAGAGACATTTCTTCAACTTGTCTCTGAAGGAAT 919
861
843 CCAAGACATTTTAAGTGAACCTACAGAGACATTTCTTCAACTTGTCTCTGAAGGAAT 902
844
920 GCTGTTCTCAGTGGGTAACTTCTCCGAGACACTGAGTTGTTTTCAGGCGCA 979
921
903 GCTGTTCTCAGTGGGTAACTTCTCCGAGACACTGAGTTGTTTTCAGGCGCA 962
904
980 TGAGCATGTTTTCGAAACCCATCACATCTCTGAATTTACTTGAATAGTTATCCCAAGAC 1039
981
963 TGAGCATGTTTTCGAAACCCATCACATCTCTGAATTTACTTGAATAGTTATCCCAAGAC 1022
964
1040 CAAAGCCAAAGATCTTTCATAGTGTATGATCCCAATGCTGGACCTCATCGGGATAA 1099
1041
1023 CAAAGCCAAAGATCTTTCATAGTGTATGATCCCAATGCTGGACCTCATCGGGATAA 1082
1024
1100 GAAGCTGAAGTACTACATCCAGAGTTGTTTACTCTGGCTCTACCCACCTCATCGAGG 1159
1101
1083 GAAGCTGAAGTACTACATCCAGAGTTGTTTACTCTGGCTCTACCCACCTCATCGAGG 1142
1084
1160 GGGAGGGGGTCTCTACTCCGCGCACCTGGCCCTGAGGCTGTACCATATCACTGACCA 1219
1161
1143 GGGAGGGGGTCTCTACTCCGCGCACCTGGCCCTGAGGCTGTACCATATCACTGACCA 1202
1144
1220 GGTCCATCTTACCCCATATGATGAGTTTACTGGAATGTGCTTTCAGAACTTCGCGCT 1279
1221
1203 GGTCCATCTTACCCCATATGATGAGTTTACTGGAATGTGCTTTCAGAACTTCGCGCT 1262
1204
1280 GGTCCAGAGAAACAAAGGCTTCAGACATTTGATATCGAGAGAGAAACAAATAA 1339
1281
1263 GGTCCAGAGAAACAAAGGCTTCAGACATTTGATATCGAGAGAGAAACAAATAA 1322
1264

QY 1340 CATCTGCTCCTATGTAGATCTGATGTTAGTACATAGTAAACCTCAAGAGATGATTGA 1399
1341
DB 1323 CATCTGCTCCTATGTAGATCTGATGTTAGTACATAGTAAACCTCAAGAGATGATTGA 1382
1324
QY 1400 TATTTGGTCTCAGTTGCGAGAGTGCCTATTAAATGCTAAATAGATACAAACTCAATT 1459
1401
DB 1383 TATTTGGTCTCAGTTGCGAGAGTGCCTATTAAATGCTAAATAGATACAAACTCAATT 1442
1384
QY 1460 KGSATWGAAGGGTWTGTTGATWGGYCCCATGTTGGGGTCTCACATTAGAGTAATTC 1519
1461
DB 1443 TGCATAGAAGGTGATTTTGAATAGTTCCCATGTTGTTCTCATAGTAATTC 1502
1444
QY 1520 TATTTNAAC--ATCAAAATTTGCTTTATGAGTGATACCATTT-ANGGCTCTAANCCT 1576
1521
DB 1503 TGTATTAACCATGAATTTGCTTTATGAGTGATACCATTTGAGGCGCTCTAANCCT 1562
1504
QY 1577 ---CATTTGNACTCAGTGAAGAGGAAAGCGGAGAGGTAAATTTTATGATGTA 1631
1578
DB 1563 TCAATTGCTACTCAGTGAAG-AGGAAAGCGGAAGATGGTAATTTTTTTACGGA 1619
1564
RESULT 4
BD093191
LOCUS
DEFINITION
Useful polypeptide.
ACCESSION
BD093191
VERSION
BD093191.1 GI:22638779
KEYWORDS
WO 0100848-A/1.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 1912)
AUTHORS
Sasaki,K., Shiraishi,N., Natsume,A., Yamada,Y., Nakagawa,S. and Sekine,S.
TITLE
Useful polypeptide
JOURNAL
Patent: WO 0100848-A 1 04-JAN-2001;
KYOKA HAKKO KOGYO CO LTD,KATSUTOSHI SASAKI,NORIHIKO SHIRAISHI,
AYUMI NATSUME,YOJI YAMADA,SATOSHI NAKAGAWA,SUSUMU SEKINE
COMMENT
OS Homo sapiens (human)
PN WO 0100848-A/1
PD 04-JAN-2001
PF 29-JUN-2000 WO 2000JP004304
PR 29-JUN-1999 JP 99P 183437,16-MAR-2000 JP 00P 074757 P1
KATSUTOSHI SASAKI,NORIHIKO SHIRAISHI,AYUMI
NATSUME,YOJI YAMADA,
PI SATOSHI NAKAGAWA,SUSUMU SEKINE
PC C12N15/54,C12N9/10,C12N5/10,C12N1/21,A61K48/00,A61K45/00, PC
A61K39/395,
PC A61K35/00,A61K31/711,A61P35/00,A61P29/00,A01K67/027,A01H5/00,
PC G01N33/53,
PC G01N33/15,G01N33/05
CC
FH Key Location/Qualifiers.
FEATURES
source
1..1912
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 518 a 430 c 457 g 507 t
ORIGIN
Query Match 88.1%; Score 1504.2; DB 6; Length 1912;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1582; Conservative 8; Mismatches 27; Indels 10; Gaps 6;
QY 11 GCAGCGGAGCGGACGAGCGGCAACAGTGCAGGAGGTAGCAGAGCCGAGCA 70
12
DB 2 GCAGCGGAGCGGACGAGCGGCAACAGTGCAGGAGGTAGCAGAGCCGAGCA 59
3
QY 71 GTCCCTCCGCGGACACCGCGGCGCGCTCCGCGGCGCGCATGGAGCGTGAGCT 130
72

Db 60 GTCCCTGCCGCGACACCGCGCGCGCTCCGGGCGCGCGCATGGAGCGGTGAGCT 119
Qy 131 GCGGGGTCGCGGGGCTGAGCGCGCGGAGCGCGCGGACGTGGATGTGGCGCGATCTC 190
Db 120 GCGGGGTCGCGGGGCTGAGCGCGCGGAGC--GCGGGACGTGGATGTGGCGCGATCTC 178
Qy 191 CGGCCCTTCCGCCCGCGCGGAGCTGGAGCTGCTCCGACACAGATATGAGAATGAG 250
Db 179 CGGCCCTTCCGCCCGCGCGGAGCTGGAGCTGCTCCGACACAGATATGAGAA:GAG 238
Qy 251 TGTGGACGTCGAAGAATTAAGTGTGTGGGTAICCTGATGATGGAATGTCTTCAITTA 310
Db 239 TGTGGACGTCGAAGAATTAAGTGTGTGGTATCTGATGATGCAATGTCTTCAITTA 298
Qy 311 TTTTATATGGAAGTCTCCAAAGCAGTAGCCCAAGAAAAATGAAAAAGGGGAAGTAT 370
Db 299 TTTTATATGGAAGTCTCCAAAGCAGTAGCCCAAGAAAAATGAAAAAGGGGAAGTAT 356
Qy 371 AATACCAAGAGAGTCTCGAAGATATCTACCCCTCCGAGGCATCTGGAACCGAGA 430
Db 359 AATACCAAGAGAGTCTCGAAGATATCTACCCCTCCGAGGCATCTGGAACCGAGA 418
Qy 431 GCAAGAGAGTGAACCGCAGTAGAACCCCATCTGAGCATGTGACCAACGACGCGG 490
Db 419 GCAAGAGAGTGAACCGCAGTAGAACCCCATCTGAGCATGTGACCAACGACGCGG 478
Qy 491 GGAGGCGGCGAGCTCTCCAATATAGCCATCTGAACCTCTGAGCATCTGAGCGGT 550
Db 479 GGAGGCGGCGAGCTCTCCAATATAGCCATCTGAACCTCTGAGCATCTGAGCGGT 538
Qy 551 CACGTCGGTGTGACGGGTTTAAACACTTGGCGACAGATTTAAAGACTTTCTCTCTA 610
Db 539 CACGTCGGTGTGACGGGTTTAAACACTTGGCGACAGATTTAAAGACTTTCTCTCTA 598
Qy 611 TTGAGATCCGCAATATTCTACTGCTTATAGATCAGCGGGATAGTGTGCAAGAAC 670
Db 599 TTGAGATCCGCAATATTCTACTGCTTATAGATCAGCGGGATAGTGTGCAAGAAC 658
Qy 671 TTTCTTGTGCTGGCGATTAACTGCTCCTACTCCACATTTTGCAGAGCAAGCAATCG 730
Db 659 TTTCTTGTGCTGGCGATTAACTGCTCCTACTCCACATTTTGCAGAGCAAGCAATCG 718
Qy 731 GGATCCTGGGCGCAAGAACGACCGAGGACCAACAGGTGGTGGAGCTTCTCTCT 790
Db 719 GGAATCCTGGGCGCAAGAACGACCGAGGACCAACAGGTGGTGGAGCTTCTCTCT 778
Qy 791 GGCCAGACACCCCGAGAGACACCCAGCCCTTCAGATATGCTGAAATTTGAGAG 850
Db 779 GGCCAGACACCCCGAGAGACACCCAGCCCTTCAGATATGCTGAAATTTGAGAG 838
Qy 851 TGAGAGCACCAAGACATTTATGTGGAACCTACAGACACATTTCTTCAACTTGTCTCT 910
Db 839 TGAGAGCACCAAGACATTTATGTGGAACCTACAGACACATTTCTTCAACTTGTCTCT 898
Qy 911 GAGGAGTGTCTTTTCTCAGTGGGTAGTACTTCTCCGCGACACACTGAGTTTGT 970
Db 899 GAGGAGTGTCTTTTCTCAGTGGGTAGTACTTCTCCGCGACACACTGAGTTTGT 958
Qy 971 CAAGGCGGATGACGATGTTTGTGAACACCCATCATCCTGAAATTTACTTGAATAGTT 1030
Db 959 CAAGGCGGATGACGATGTTTGTGAACACCCATCATCCTGAAATTTACTTGAATAGTT 1018
Qy 1031 ATCAAGACCAAGACCAAGATCTTTATAGTGAATGATGCCAATAGCTGGACCTCA 1090
Db 1019 ATCAAGACCAAGACCAAGATCTTTATAGTGAATGATGCCAATAGCTGGACCTCA 1078
Qy 1091 TCGGGATAAGAGCTGAGTACTACATCCGAGAGTTGTTTACTTGGGCTCTACCCACC 1150
Db 1079 TCGGGATAAGAGCTGAGTACTACATCCGAGAGTTGTTTACTTGGGCTCTACCCACC 1138
Qy 1151 CTATGAGGGGAGGGGGTTCCTCTACTCCGCCACCTGGCCCTGAGGCTGTACCATAT 1210
Db 1139 CTATGAGGGGAGGGGGTTCCTCTACTCCGCCACCTGGCCCTGAGGCTGTACCATAT 1198

Qy 1211 CACTGACCAGTCCATCTACCCCATTTGATGACGTTTATACTGGAATGTCCTTCAGAA 1270
Db 1199 CACTGACCAGTCCATCTACCCCATTTGATGACGTTTATACTGGAATGTCCTTCAGAA 1258
Qy 1271 ACTGGGCTCGTTCCAGAGAAACACAAAGGCTTCAGACATTTGATATCGAGGAGAAAA 1330
Db 1259 ACTGGGCTCGTTCCAGAGAAACACAAAGGCTTCAGACATTTGATATCGAGGAGAAAA 1318
Qy 1331 CAAAAATAACATCTGCTCCTATGTAGATCTGTAGTCTGTAGTACATAGTAAAAACCTCAAGA 1390
Db 1319 CAAAAATAACATCTGCTCCTATGTAGATCTGTAGTCTGTAGTACATAGTAAAAACCTCAAGA 1378
Qy 1391 GATGATTGATATTTGGTCTCAGTTCAGTTCAGTTCATTTAAATGCTTAAATATGATACAA 1450
Db 1379 GATGATTGATATTTGGTCTCAGTTCAGTTCAGTTCATTTAAATGCTTAAATATGATACAA 1438
Qy 1451 ACTCAATTTKGSATWRAAGGGTWTTTTGTATWGSYCCCATGTTGGGTCCTCACATTAG 1510
Db 1439 ACTCAATTTGCATAGAAAGGTGATTTTGAATAGTTCCTCATGTGTCTCACATTAG 1498
Qy 1511 AGTAATTTCTATTNAANC--ATGAAATTCCTTTATGATGATACCCATTT-ANGGCT 1567
Db 1499 AGTAATTTCTATTAAACCATGAAATTCCTTTATGATGATACCCATTTGAGGCT 1558
Qy 1568 CTAANCCCTT---CATTTGNACTCAGTGAAGAGGAAAGCGGAGAGGTAATTTTNTTT 1624
Db 1559 CTAANCCCTTCAATTTGCTACTCAGCTGAAG-AGGGAAGCGGAAGATGCTAATTTTTT 1617
Qy 1625 ATGGTGA 1631
Db 1618 TTATGGA 1624

RESULT 5
AB049584
LOCUS Homo sapiens mRNA for beta-1,3-N-acetylglucosaminyltransferase 1912 bp mRNA linear PRI 31-JAN-2001
DEFINITION bGnt-2, complete cds.
ACCESSION AB049584
VERSION AB049584.1 GI:12619293
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (sites)
AUTHORS Shiraishi,N., Natsume,A., Togayachi,A., Endo,T., Akashima,T., Yamada,Y., Imai,N., Nakagawa,S., Koizumi,S., Sekine,S., Narimatsu,H. and Sakaki,K.
TITLE Identification and characterization of three novel beta 1,3-N-acetylglucosaminyltransferases structurally related to the beta 1,3-galactosyltransferase family
J. Biol. Chem. 276 (5), 3498-3507 (2001)
JOURNAL MEDLINE 21264825
MEDLINE 11042166
PERMED
REFERENCE 2 (bases 1 to 1912)
AUTHORS Sasaki,K., Natsume,A. and Shiraishi,N.
TITLE Direct Submission
JOURNAL Submitted (02-OCT-2000) Katsutoshi Sasaki, The Tokyo Research Laboratories, Kyowa Hakko Kogyo Co., Ltd., Department of Molecular Genetics, 3-6-6 Asahi-machi, Machida-shi, Tokyo 194-8533, Japan (E-mail:ksasaki@kyowa.co.jp, Tel:81-42-725-2555(ex.2150), Fax:81-42-726-8330)
FEATURES
Location/Qualifiers
1. 1912
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..1912
/gene="bGnt-2"
234..1427
/gene="bGnt-2"

JOURNAL
99063792
MEDLINE
PUBMED
REFERENCE
2 (bases 1 to 99088)
Holmes, A. and Cotton, M.
The sequence of Homo sapiens BAC clone RP11-93M19
Unpublished (2001)
REFERENCE
3 (bases 1 to 99088)
Waterston, R. H.
Direct Submission
Submitted (22-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 99088)
Waterston, R. H.
Direct Submission
Submitted (20-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 99088)
Waterston, R. H.
Direct Submission
Submitted (29-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 99088)
Waterston, R. H.
Direct Submission
Submitted (16-APR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 29, 2002 this sequence version replaced gi:19424712.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0093M19

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPC111 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-270B14, 2000 bp overlap;
the clone sequenced to the right is RP11-642B6, 2000 bp overlap.
Actual end of this clone is at base position 38298 of RP11-642B6.

Sequence derived from one M13 subclone, base position 30875 to 30886.

Polymorphisms have been identified between AC018462, AC093159 and AC093401.

FEATURES	Location/Qualifiers
SOURCE	1..99088
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-93M19"
	/clone_lib="RPC111"
repeat_region	218..337
	/rpt_family="MIR"
repeat_region	906..1160
	/rpt_family="Alu"
repeat_region	1379..1925
	/rpt_family="ERV1"
repeat_region	2261..2297
	/rpt_family="MIR"
repeat_region	2571..2939
	/rpt_family="L1"
repeat_region	2974..3081
	/rpt_family="MIR"
repeat_region	3123..3225
	/rpt_family="Alu"
repeat_region	3486..3708
	/rpt_family="MIR"
misc_feature	3789..4320
	/note="match to EST BG721682 (NID:g14000869)"
misc_feature	3829..4087
	/note="match to EST AA377082 (NID:g2029630)"
misc_feature	3869..4691
	/note="match to EST AU137528 (NID:g10998067)"
misc_feature	4100..4563
	/note="match to EST H78875 (NID:g1056964) yu09g06.rl"
misc_feature	4141..4441
	/note="match to EST AA263148 (NID:g1898954)"
misc_feature	4142..4793
	/note="match to EST BG741357 (NID:g14052010)"
misc_feature	4180..4678
	/note="similar to Mus musculus EST AA204363 (NID:g18001007) mu30ell.rl"
misc_feature	4264..4724
	/note="match to EST AW958548 (NID:g8148232)"
misc_feature	4426..5325
	/note="similar to EST BM557603 (NID:g18799721)"
misc_feature	4468..4938
	/note="similar to Mus musculus EST BE632125 (NID:g9914895) uu57d09.y1"
misc_feature	4498..4879
	/note="match to EST H47990 (NID:g924042) yp75d11.rl"
misc_feature	4624..5231
	/note="match to EST BF001654 (NID:g10701929)"
misc_feature	4666..5089
	/note="match to EST AV682596 (NID:g10284459)"
misc_feature	4680..5391
	/note="match to EST BE622082 (NID:g98930322)"
misc_feature	4915..5257
	/note="match to EST R31722 (NID:g787565) yhe3d09.rl"
misc_feature	4945..5390
	/note="match to EST H13125 (NID:g877945) yj06e09.rl"
misc_feature	4948..5597
	/note="match to EST BF131058 (NID:g10970087)"
misc_feature	5026
	/note="match to EST R74552 (NID:g848922) yi58a02.rl"
misc_feature	5031..5682
	/note="match to EST BE564288 (NID:g9808008)"
misc_feature	5061..5553
	/note="match to EST N58174 (NID:g1202064) yv66a07.s1"

Db 1242 GTAGAAAACCTCAGAGATGATTGATATTGGTCTCAGITGCAGAGTGCTCATTTAAAT 1301

Qy 1435 GCTAAATAGATACAACTCAATTTKGSATWGAAGGGGTWTTTIGRATWGGYCCCATGT 1494

Db 1302 GCTAAATAGATACAACTCAATTTTGCATAGAAAGGTGTATTTTGAATAGTCCCATGT 1361

Qy 1495 TGGGGTCTACATAGAGTAATTTCTATTNANC--ATGAATTCCTTTATGAGTAT 1552

Db 1362 TGTGTTCTCATTAGAGTAATTTCTGATTAACCATGAAATTCCTTTATGAGTAT 1421

Qy 1553 ACCCATTT-ANGGCTCTAANCCTT---CATTTGNACTCAGCTGAAGAAGGGAACCGG 1608

Db 1422 ACCATTTGAGGGCTCTAAACCTTCAATTTGGTACTCAGCTGAAG-AGGGAACCGGA 1480

Qy 1609 AGAAGTAATTTTWTATGTGTA 1631

Db 1481 AGATGTAATTTTTTTTACGGA 1503

RESULT 8

AF288209

LOCUS 2516 bp mRNA linear PRI 03-AUG-2000

DEFINITION Homo sapiens beta galactosyltransferase bGalT7 mRNA, complete cds.

ACCESSION AF288209

VERSION AF288209.1 GI:9664888

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2516)

AUTHORS Gromova,I., Gromov,P. and Celis,J.

TITLE A novel member of beta-1,3-galactosyltransferase family is down regulated during bladder TCC progression

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2516)

AUTHORS Gromova,I., Gromov,P. and Celis,J.

TITLE Direct Submission

JOURNAL Submitted (18-JUL-2000) Medical Biochemistry, University of Aarhus, Ole Worms Alle 170, Aarhus C-8000, Denmark

FEATURES

source Location/Qualifiers

1..2516

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/tissue_type="bladder cancer"

43..1224

/codon_start=1

/product="beta galactosyltransferase bGalT7"

/protein_id="AAF97254.1"

/db_xref="GI:9664889"

CDs

745 a 440 c 522 g 809 t

Query Match 74.4% Score 1269.8; DB 9; Length 2516;

Best Local Similarity 97.0%; Pred. No. 2e-284;

Matches 1324; Conservative 8; Mismatches 26; Indels 7; Gaps 4;

Qy 273 TTGTTGGGTATCTGATGATGGCAATGTCTTCATTATTTTATGGAAGTCTCCAA 332

Db 58 TTGTTGGGTATCTGATGATGGCAATGTCTTCATTATTTTATGGAAGTCTCCAA 117

Qy 333 AGCAGTAGCCCAAGAAAATGGAAGGGGAAGTAATATATCCCAAGAGAAGTCTCG 392

Db 118 AGCAGTAGCCCAAGAAAATGGAAGGGGAAGTAATATATCCCAAGAGAAGTCTCG 177

Qy 393 AAGATATATACCCCTCCGAGGCGATACTGGAACCGAGAGCAAGAGCTGAACCGCGAG 452

Db 178 AAGATATATACCCCTCCGAGGCGATACTGGAACCGAGAGCAAGAGCTGAACCGCGAG 237

Qy 453 TACAAACCCATCTCTGAGCATGCTGACCAACAGAGCGGGAGGGGCGGAGCTCTCCAAT 512

Db 238 TACAAACCCATCTCTGAGCATGCTGACCAACAGAGCGGGAGGGGCGGAGCTCTCCAAT 297

Qy 513 ATAAGCCATCTGAATCTGGAACCTGACCTGAGGGTCACTGCTGGTGGTGTACGGGTTT 572

Db 298 ATAAGCCATCTGAATCTGGAACCTGACCTGAGGGTCACTGCTGGTGGTGTACGGGTTT 357

Qy 573 AACAACTTGGCGGACAGATTTAAAGACTTCTGCTGTATTTGAGATGCCGCAATATTCGA 632

Db 358 AACAACTTGGCGGACAGATTTAAAGACTTCTGCTGTATTTGAGATGCCGCAATATTCGA 417

Qy 633 CTGCTTATAGATCAGCGGATTAAGTGTCAAAGAAACCTTTCTTGTCTGGCGATTAAAG 692

Db 418 CTGCTTATAGATCAGCGGATTAAGTGTCAAAGAAACCTTTCTTGTCTGGCGATTAAAG 477

Qy 693 TCCCTCACTCCACATTTTGGCAGAGGAAGCAATCCGGGAATCCTGGGGCCAAAGAAAGC 752

Db 478 TCCCTCACTCCACATTTTGGCAGAGGAAGCAATCCGGGAATCCTGGGGCCAAAGAAAGC 537

Qy 753 AACGAGGGAACCAACCGTGTGGAGTCTTCTGCTGGGCCAGACACCCGCCAGAGAC 812

Db 538 AACGAGGGAACCAACCGTGTGGAGTCTTCTGCTGGGCCAGACACCCGCCAGAGAC 597

Qy 813 AACCAACCCGACCTTTCAGATATGCTGAAATTTGAGAGTGAGAAGCACCAAGACATCTT 872

Db 598 AACCAACCCGACCTTTCAGATATGCTGAAATTTGAGAGTGAGAAGCACCAAGACATCTT 657

Qy 873 ATGTGGAAGTACAGAGACACTTCTTCAACTGTCTCTGAAGGAGTGTCTTTCTCAGG 932

Db 658 ATGTGGAAGTACAGAGACACTTCTTCAACTGTCTCTGAAGGAGTGTCTTTCTCAGG 717

Qy 933 TGGGTAAGTACTTCTGCCAGACACTGAGTTTGTTCATGAGGGGATGACGATGTTT 992

Db 718 TGGGTAAGTACTTCTGCCAGACACTGAGTTTGTTCATGAGGGGATGACGATGTTT 777

Qy 993 GTGAACACCCATCAGATCCTGAAATTTAGTATTTATCCAGACCAAGACCAAGAT 1052

Db 778 GTGAACACCCATCAGATCCTGAAATTTAGTATTTATCCAGACCAAGACCAAGAT 837

Qy 1053 CTCCTCATAGTGTGATCCACAATGCTGACCTCATCGGATAGCAAGCTGAGTAC 1112

Db 838 CTCCTCATAGTGTGATCCACAATGCTGACCTCATCGGATAGCAAGCTGAGTAC 897

Qy 1113 TACATCCCAAGAGTTGTTTACTCTGGCCTCTACCCACCTATGACAGGGGAGGGGGTTC 1172

Db 898 TACATCCCAAGAGTTGTTTACTCTGGCCTCTACCCACCTATGACAGGGGAGGGGGTTC 957

Qy 1173 CTCCTACTCGGCCACTCGCCTGAGGCTGTACCATATCTACTGACAGGTCTCATCTAC 1232

Db 958 CTCCTACTCGGCCACTCGCCTGAGGCTGTACCATATCTACTGACAGGTCTCATCTAC 1017

Qy 1233 CCATTGTAGCTGTTTATATGGAATGTGCCTTTCAGAACTCGGCCTCGTTCAGAGAAA 1292

Db 1018 CCATTGTAGCTGTTTATATGGAATGTGCCTTTCAGAACTCGGCCTCGTTCAGAGAAA 1077

Qy 1293 CACAAAGGCTTCAGGACATTTGATATCGAGGAGAAAAACAAAAATACATCTGCTCCTAT 1352

Db 1078 CACAAAGGCTTCAGGACATTTGATATCGAGGAGAAAAACAAAAATACATCTGCTCCTAT 1137

Qy 1353 GTAGATCTGATGTTAGTACATAGTAGAAAACCTCAAGAGATGATGATATTGGTCTCAG 1412

Db 1138 GTAGATCTGATGTTAGTACATAGTAGAAAACCTCAAGAGATGATGATATTGGTCTCAG 1197

Qy 1413 TTGCAGAGTGCTCATTTTAAATGCTTAAATAGATACAACTCAATTTKGSATWGAAGG 1472

Db 1198 TTGCAGAGTGCTCATTTTAAATGCTTAAATAGATACAACTCAATTTTGCATAGAAAGT 1257

Qy	1473	GI:27862722	CGCTGAAG	AGGGAAGCGGGAAGTGAATTTTATGGTGA	1631
Db	1258	GI:27862722	CGCTGAAG	AGGGAAGCGGGAAGTGAATTTTATGGTGA	1421
Qy	1531	GI:27862722	CGCTGAAG	AGGGAAGCGGGAAGTGAATTTTATGGTGA	1631
Db	1318	GI:27862722	CGCTGAAG	AGGGAAGCGGGAAGTGAATTTTATGGTGA	1421
Qy	1587	GI:27862722	CGCTGAAG	AGGGAAGCGGGAAGTGAATTTTATGGTGA	1631
Db	1378	GI:27862722	CGCTGAAG	AGGGAAGCGGGAAGTGAATTTTATGGTGA	1421
RESULT 9					
LOCUS	BD156964		2442 bp	DNA	linear
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.				
ACCESSION	BD156964				
VERSION	BD156964.1	GI:27862722			
KEYWORDS	JP 2002191363-A/11807				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
AUTHORS	Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.				
TITLE	Primer for synthesizing full-length cDNA and use thereof				
JOURNAL	Patent: JP 2002191363-A/11807 09-JUL-2002; HELIX RESEARCH INSTITUTE				
COMMENT	OS Homo sapiens (human) PN JP 2002191363-A/11807 PD 09-JUL-2002 PF 28-JUL-2000 JP 2000280990 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SAITO, PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU, PI KEIICHI NAGAI, TETSUJI OTSUKI PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC 10, C12P21/02,C12P1/08,C06F17/30,C12N15/00,C12N5/00 CC Primer for synthesizing full-length cDNA and use thereof FH Key Location/Qualifiers FT CDS Location/Qualifiers 1..2442 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"				
BASE COUNT	717 a	433 c	503 g	789 t	
ORIGIN					
Query Match	72.18;	Score 1231.4;	DB 6;	Length 2442;	
Best Local Similarity	97.08;	Pred. No. 1.7e-275;			
Matches 1289;	Conservative 8;	Mismatches 25;	Indels 7;	Gaps 4;	
Qy	313	TTATTATGGAAGTCTCCAAAGCAGTACCCCAAGGGAAGTGAATTTTATGGTGA	372		
Db	1	TTATTATGGAAGTCTCCAAAGCAGTACCCCAAGGGAAGTGAATTTTATGGTGA	60		
Qy	373	TACCAAGAGAGTCTCGAAGATATCTACCCCTCCCGAGGCATCTGGAACCGGAGC	432		
Db	61	TACCAAGAGAGTCTCGAAGATATCTACCCCTCCCGAGGCATCTGGAACCGGAGC	120		
Qy	433	AAGGAGAGCTGAGCGGAGTACCAACCCCATCTGAGCATGCTGACCAACGACGCGGG	492		
Db	121	AAGGAGAGCTGAGCGGAGTACCAACCCCATCTGAGCATGCTGACCAACGACGCGGG	180		
Qy	493	AGCGGGCAGGCTCTCCAAATATAGGCATCTGAACTACTCGCACTGACCTGAGGGTCA	552		
Db	181	AGCGGGCAGGCTCTCCAAATATAGGCATCTGAACTACTCGCACTGACCTGAGGGTCA	240		

DB	1320	ACGG	1324	QY	433	AAAGAGAGCTGAACCGGAGTACACCCCATCTCTGAGCATCTCTGAGCAGGAGG 492
				DB	121	AAGAGAGAGCTGAACCGGAGTACACCCCATCTCTGAGCATCTCTGAGCAGGAGG 180
				QY	493	AGCGGGCAGGCTCTCCAATATAGGCATCTGAAGTCTGCGAAGCTGACCTGAGGCTCA 552
				DB	181	AGCGGGCAGGCTCTCCAATATAGGCATCTGAAGTCTGCGAAGCTGACCTGAGGCTCA 240
				QY	553	CGTCGGTGGTACGGGTTTAAACAACCTGCGGACAGATTTAAAGACTTTCTGCTGTATT 612
				DB	241	CGTCGGTGGTACGGGTTTAAACAACCTGCGGACAGATTTAAAGACTTTCTGCTGTATT 300
				QY	613	TGAGATGCCGCAATATTCACCTGCTTATAGATCAGCGGATAGTGTGCAAGAAGACCTT 672
				DB	301	TGAGATGCCGCAATATTCACCTGCTTATAGATCAGCGGATAGTGTGCAAGAAGACCTT 360
				QY	673	TCTTGTGTGGGATTAAAGTCCCTCACTCCACATTTTCCAGAGGCAAGCAATCCGGG 732
				DB	361	TCTTGTGTGGGATTAAAGTCCCTCACTCCACATTTTCCAGAGGCAAGCAATCCGGG 420
				QY	733	AATCCTGGGGCCAAAGAAAGCAAGCGGAGGAAACCAAGCGTGGTGGAGTCTTCTGCTGG 792
				DB	421	AATCCTGGGGCCAAAGAAAGCAAGCGGAGGAAACCAAGCGTGGTGGAGTCTTCTGCTGG 480
				QY	793	GCCAGACACCCCGGAGGACCAACCCCGGACCTTTTCAGATATGCTGAAATTTGAGAGTG 852
				DB	481	GCCAGACACCCCGGAGGACCAACCCCGGACCTTTTCAGATATGCTGAAATTTGAGAGTG 540
				QY	853	AGAAGCACCAGACATCTTATGTGGAATACAGAGACACATTTCTTCAACITGTCTCTGA 912
				DB	541	AGAAGCACCAGACATCTTATGTGGAATACAGAGACACATTTCTTCAACITGTCTCTGA 600
				QY	913	AGAAAGTGTCTTCTCAGTGGGTAAAGTACTTCTCGCCAGACACTGAGTGTGTTTCA 972
				DB	601	AGAAAGTGTCTTCTCAGTGGGTAAAGTACTTCTCGCCAGACACTGAGTGTGTTTCA 660
				QY	973	AGGGCGATGACGATGTTTTTGTGAACACCCCATCATCTCTGAAATTTACITGAAATGTTAT 1032
				DB	661	AGGGCGATGACGATGTTTTTGTGAACACCCCATCATCTCTGAAATTTACITGAAATGTTAT 720
				QY	1033	CCAGACCAAAGCCAAAGATCTCTTATAGTGAITGTATCCACATGCTGGACCTCATC 1092
				DB	721	CCAGACCAAAGCCAAAGATCTCTTATAGTGAITGTATCCACATGCTGGACCTCATC 780
				QY	1093	GGATTAAGAGCTGAAGTACTACATCCAGAGTCTTACTCTGGCTCTACCCACCT 1152
				DB	781	GGATTAAGAGCTGAAGTACTACATCCAGAGTCTTACTCTGGCTCTACCCACCT 840
				QY	1153	ATCAGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGCTGTACCATATCA 1212
				DB	841	ATCAGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGCTGTACCATATCA 900
				QY	1213	CTGACCAAGTCCATCTCTACCCCATCTGATGACCTTTTACTGGAATGTGCCITCAGAAC 1272
				DB	901	CTGACCAAGTCCATCTCTACCCCATCTGATGACCTTTTACTGGAATGTGCCITCAGAAC 960
				QY	1273	TCGGCTCTGTTCCAGAGAAACAAAGGCTTCAGGACATTTGATATCGAGGAGAAAAACA 1332
				DB	961	TCGGCTCTGTTCCAGAGAAACAAAGGCTTCAGGACATTTGATATCGAGGAGAAAAACA 1020
				QY	1333	AAAATAACATCTCTCTCTATGTAGATCTGTATGTAGTACATAGTAAACCTCAAGAGA 1392
				DB	1021	AAAATAACATCTCTCTCTATGTAGATCTGTATGTAGTACATAGTAAACCTCAAGAGA 1080
				QY	1393	TGATTTGATTTGGTCTCAGTTCAGAGTCTCATTTTAAATGCTTAAATAGATACAAAC 1452
				DB	1081	TGATTTGATTTGGTCTCAGTTCAGAGTCTCATTTTAAATGCTTAAATAGATACAAAC 1140
				QY	1453	TCATTTTGSATWCRAGGGTWTTTTGTATGAGTCCATGTTGGGCTCTCACAATTAGAG 1512
				DB	1141	TCAATTTTGCATGAGAAAGGTGATTTTGAATAGTTCCTGTTGTTCTTCTCACAATTAGAG 1200
				QY	1513	TAATTTCTATTNNAAC--ATGAAATTTGCCCTTTTATGAGTGATACCCATTTT-ANGGCCTCT 1569
DB	1320	ACGG	1324			
RESULT 10						
AK002009						
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						
FEATURES						
source						
CDS						
BASE COUNT						
ORIGIN						
Query Match						
Best Local Similarity						
Matches 1285; Conservative						
QY	313	TTATTATGGAAGTCTCCAAAAGCAGTACGCCAAGAAAAAATGGAAGGGGAAAGTAATAA 372				
DB	1	TTATTATGGAAGTCTCCAAAAGCAGTACGCCAAGAAAAAATGGAAGGGGAAAGTAATAA 60				
QY	373	TACCAAGAGAGAGTCTTGGAGATATCTACCCCTCCCGAGGCATCTGGAACCGAGAGC 432				
DB	61	TACCAAGAGAGAGTCTTGGAGATATCTACCCCTCCCGAGGCATCTGGAACCGAGAGC 120				

```

||||| 1201 TAATTCGTATTAAACCATGAAATGCGTTTATGAGTGATACCCATTGAGGCGCTCT 1260
||||| 1570 AANCCCTT---CATTTGACTCAGTGAAGGAAAGCGGAGAGGTAATTTTAT 1626
||||| 1261 AACCCTTCAATTTGTTACTCAGTGAAG-AGGAAAGCGGAAGATGGTAATTTTTTT 1319
|| 1627 GGTGA 1631
||
1320 ACGGA 1324

RESULT 11
HSA6077
LOCUS Homo sapiens beta3gal-T5 gene. 1197 bp mRNA linear PRI 11-MAY-2000
DEFINITION Homo sapiens beta3gal-T5 gene.
ACCESSION AJ006077
VERSION AJ006077.1 GI:7799920
KEYWORDS beta-1,3-galactosyltransferase; beta3gal-T5 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Anado,M., Carneiro,F. and Clausen,H.
TITLE Cloning and expression of two beta-1,3-galactosyltransferases:
beta3gal-T5 and beta3gal-T6
JOURNAL Unpublished
AUTHORS Anado,M.
TITLE Direct Submission
JOURNAL Submitted (11-MAY-1998); Amado M., Department of Oral Diagnostics,
Royal dental School, Nørre Alle 20, 2200 Copenhagen, DENMARK
FEATURES
source 1..1197
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..1197
/gene="beta3gal-T5"
1..1197
/codon_start=1
/product="beta-1,3-galactosyltransferase"
/protein_id="CA991546.1"
/db_xref="GI:7799921"
/translation="MSVGRRRRIKLLILMNAVFIYFINEVSKSSQEKNGKEVII
PKKFKWISTPPPEAYNREDEKLNROYNPILSLTNOTGEAGRLSNISHLNCEPDLR
VTSVVTGFNNLPDRFDLLYLRCBNVSLIDOPDKCAKPELLLAIKSLIHPFARRQ
AIRSAGQESNAGNQTIVRVFLGQTPPEDNHPDLSMDLKFSEKHKQDILMNNYRDIF
FNLSLKEVLFLRWYSTSCPDETFVFKGDDDFVFNTHILNLSLSTKAKDLFIQDV
IHNAGPHRDKLKYIPEVYVYSG.YPPYAGGGGLYSGHLALRLYHITDQVHLPIQD
VYTGMLQKLGLVPEKHKGFRTEIDIEKNKNICSYVDLMLVHSRKPQEMDIWSOLO
SAHLKC"
BASE COUNT 347 a 274 c 281 g 295 t
ORIGIN

Query Match 69.2%; Score 1181; DB 9; Length-1197;
Best Local Similarity 99.7%; Pred. No. 7.9e-264;
Matches 1194; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 246 ATGAGTGTTCGAGCTCGAAGAAATAA---CTTGTGGGTATCCTGATGATGCGAATGTC 302
||||| 1 ATGAGTGTTCGAGCTCGAAGAAATAAAGTTCTTGTGGGTATCCTGATGATGCGAATGTC 60
||||| 303 TCATATTTATTTATATGGAAGTCTCCAAAGCAGTAGCCCAAGAAATAAATGCAAGGG 362
||||| 61 TCATATTTATTTATATGGAAGTCTCCAAAGCAGTAGCCCAAGAAATAAATGCAAGGG 120
||||| 363 GAAGTAATAATACCAAGAGAAGTTCTCGAAGATATCTACCCCTCCCGAGGCATACG 422
||||| 121 GAAGTAATAATACCAAGAGAAGTTCTCGAAGATATCTACCCCTCCCGAGGCATACG 180

```

RESULT 12
AY043479

LOCUS AY043479 2672 bp mRNA linear ROD 02-SEP-2001
DEFINITION Mus musculus beta-1,3-N-acetylglicosaminyltransferase mRNA,

423 AACCCGAGCAAGAGAAGCTGAACCGGCGAGTACAACCCCATCCTGAGCATGCTGACCAAC 482
||||| 181 AACCCGAGCAAGAGAAGCTGAACCGGCGAGTACAACCCCATCCTGAGCATGCTGACCAAC 240
QY 483 CAGACGGGGAGGGGGGAGGCTCTCCATATATAAGCCATCTGAACCTACTGCGAAGCTGAC 542
Db 241 CAGACGGGGAGGGGGGAGGCTCTCCAATATAAGCCATCTGAACCTACTGCGAAGCTGAC 300
QY 543 CTGAGGTCACGTCGGTGGTTTACGGGTTTAAACAACCTTGCAGGACAGATTTAAACACTTT 602
Db 301 CTGAGGTCACGTCGGTGGTTTACGGGTTTAAACAACCTTGCAGGACAGATTTAAACACTTT 360
QY 603 CTGCTGTATTTGAGATGCGGCAATTTACTACTGCTTATAGATCAGCCGAGTAAGTGTGCA 662
Db 361 CTGCTGTATTTGAGATGCGGCAATTTACTACTGCTTATAGATCAGCCGAGTAAGTGTGCA 420
QY 663 AAGAAACCTTTCTTGTGCTGGCGATTAAAGTCCCTCACTCCACATTTTCCCAAGAGGCA 722
Db 421 AAGAAACCTTTCTTGTGCTGGCGATTAAAGTCCCTCACTCCACATTTTCCCAAGAGGCA 480
QY 723 GCAATCCGGGAATCTCTGGGGCCCAAGAAACACGAGGGAACCAACCGTGTGCGAGTC 782
Db 481 GCAATCCGGGAATCTCTGGGGCCCAAGAAACACGAGGGAACCAACCGTGTGCGAGTC 540
QY 783 TTCTCTGCTGGGCAGACACCCCGAGAGGACACACCCCGAGCTTTTCAGATATGCTGAAA 842
Db 541 TTCTCTGCTGGGCAGACACCCCGAGAGGACACACCCCGAGCTTTTCAGATATGCTGAAA 600
QY 843 TTGAGAGTGAGAAGCACCACAGACATCTTATGTGGAAGTACAGAGACACATTTCTTCAAC 902
Db 601 TTGAGAGTGAGAAGCACCACAGACATCTTATGTGGAAGTACAGAGACACATTTCTTCAAC 660
QY 903 TTGCTCTCAAGGAAGTGTCTTTCTCAGGTGGGTAAAGTACTTCTGCCACACACTGAG 962
Db 661 TTGCTCTCAAGGAAGTGTCTTTCTCAGGTGGGTAAAGTACTTCTGCCACACACTGAG 720
QY 963 TTTGTTTTCAAGGGCGATGACGATGTTTTTGTGAACACCCCATCACATCCTGAATTTG 1022
Db 721 TTTGTTTTCAAGGGCGATGACGATGTTTTTGTGAACACCCCATCACATCCTGAATTTG 780
QY 1023 AATAGTTTATCCAGACCAACCAAGCCAAAGATCTCTTCATAGGTGATGTATCCCAATGCT 1082
Db 781 AATAGTTTATCCAGACCAACCAAGCCAAAGATCTCTTCATAGGTGATGTATCCCAATGCT 840
QY 1083 GGACCTCATCGGGATAAGAAAGCTGAAGTACTACATCCCAAGAAAGTTGTTTACTCTGGCCTC 1142
Db 841 GGACCTCATCGGGATAAGAAAGCTGAAGTACTACATCCCAAGAAAGTTGTTTACTCTGGCCTC 900
QY 1143 TACCCACCTATGCGAGGGAGGGGGTTCCTCTACTCGGGCCACTGGCCCTGAGGCTG 1202
Db 901 TACCCACCTATGCGAGGGAGGGGGTTCCTCTACTCGGGCCACTGGCCCTGAGGCTG 960
QY 1203 TACCATATCACTGACGAGTGCCATCTTACCCCATTTGATGAGCTTTATAGTGGATGTGC 1262
Db 961 TACCATATCACTGACGAGTGCCATCTTACCCCATTTGATGAGCTTTATAGTGGATGTGC 1020
QY 1263 CTTCAAGAACTCGGCTCGTTCCAGAGAAACAAAGGCTTCAGGACATTTGATATCGAG 1322
Db 1021 CTTCAAGAACTCGGCTCGTTCCAGAGAAACAAAGGCTTCAGGACATTTGATATCGAG 1080
QY 1323 GAGAAACAAATAATACATCTGCTCTTATGTAGATCTGATGTTAGTACATAGTAAAA 1382
Db 1081 GAGAAACAAATAATACATCTGCTCTTATGTAGATCTGATGTTAGTACATAGTAAAA 1140
QY 1383 CTTCAAGAGATGATTTGTTGGTCTCAGTGTGAGAGTGTCTCATTTAAATGTAA 1439
Db 1141 CTTCAAGAGATGATTTGTTGGTCTCAGTGTGAGAGTGTCTCATTTAAATGTAA 1197


```

RESULT 13
AL390027/c
LOCUS
DEFINITION
  Human DNA sequence from clone RP11-368G3 on chromosome X, complete
  sequence.
ACCESSION
  AL390027
VERSION
  AL390027.1; GI:10443456
KEYWORDS
  HTG.
SOURCE
  Homo sapiens (human)
  Homo sapiens
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 200191)
  Chapman, J.
  Direct Submission
  Submitted (06-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
  Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
  humquerry@sanger.ac.uk clone requests: clonerequests@sanger.ac.uk
  On Oct 1, 2000 this sequence version replaced gi:10279715.
  During sequence assembly data is compared from overlapping clones.
  Where differences are found these are annotated as variations
  together with a note of the overlapping clone name. Note that the
  variation annotation may not be found in the sequence submission
  corresponding to the overlapping clone, as we submit sequences with
  only a small overlap as described above.
  The following abbreviations are used to associate primary accession
  numbers given in the feature table with their source databases:
  Em.; EMBL; SW.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information
  on the WORMPEP database can be found at
  http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
  was generated from part of bacterial clone contigs of human
  chromosome X, constructed by the Sanger Centre Chromosome X Mapping
  Group. Further information can be found at
  http://www.sanger.ac.uk/HGP/ChrX
  This sequence is the entire insert of clone RP11-368G3 This
  sequence was finished as follows unless otherwise noted: all
  regions were either double-stranded or sequenced with an alternate
  chemistry or covered by high quality data (i.e., phred quality >=
  30); an attempt was made to resolve all sequencing problems, such
  as compressions and repeats; all regions were covered by at least
  one plasmid subclone or more than one M13 subclone; and the
  assembly was confirmed by restriction digest. RP11-368G3 is from
  the library RPI-11.2 constructed by the group of Pieter de Jong.
  For further details see
  http://www.chori.org/bacpac/home.htm
  VECTOR: pBACE3.6.
FEATURES
  Location/Qualifiers
    1..200191
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="X"
      /clone="RP11-368G3"
      /clone_lib="RPCI-11.2"
      1..5390
        /note="L1MAL repeat: matches 780..6198 of consensus"
        repeat_region
      5395..7739
        /note="L1MB1 repeat: matches 3353..5691 of consensus"
        repeat_region
      7803..8244
        /note="L1MB1 repeat: matches 5706..6165 of consensus"
        repeat_region
      8439..8794
        /note="178 copies 2 mer aa 55% conserved"
        repeat_region
      8857..8978
        /note="61 copies 2 mer gg 61% conserved"
        repeat_region
      9316..16924
        /note="L1MAL repeat: matches 1389..6304 of consensus"
        repeat_region
      16978..17379
        /note="MSTA repeat: matches 1..422 of consensus"
        repeat_region
      17394..19014
        /note="MSTA-interval repeat: matches 1..1636 of consensus"
        repeat_region
      19015..19402
        /note="MSTA repeat: matches 1..426 of consensus"

```

```

repeat_region
22130..22169
  /note="L2 repeat: matches 2711..2750 of consensus"
  repeat_region
22707..24172
  /note="L1MB2 repeat: matches 4655..6150 of consensus"
  repeat_region
24594..24774
  /note="HALL repeat: matches 1535..1722 of consensus"
  repeat_region
24881..29302
  /note="L1PA16 repeat: matches 1855..6154 of consensus"
  repeat_region
29303..29667
  /note="THEL8 repeat: matches 1..364 of consensus"
  repeat_region
29668..30215
  /note="L1PA16 repeat: matches 1323..1855 of consensus"
  repeat_region
30216..30336
  /note="L1PB2 repeat: matches 6034..6155 of consensus"
  repeat_region
30337..30644
  /note="AluSg repeat: matches 2..297 of consensus"
  repeat_region
30645..30827
  /note="L1PB2 repeat: matches 5854..6034 of consensus"
  repeat_region
30828..31914
  /note="MERLIC repeat: matches 1..1071 of consensus"
  repeat_region
31915..32879
  /note="L1PB2 repeat: matches 4890..5859 of consensus"
  repeat_region
31915..31958
  /note="L1PA16 repeat: matches 1013..1323 of consensus"
  repeat_region
32855..35026
  /note="L1PA15-16 repeat: matches 705..1140 of consensus"
  repeat_region
35421..35872
  /note="HALL repeat: matches 674..1145 of consensus"
  repeat_region
35871..35992
  /note="HALL repeat: matches 179..298 of consensus"
  repeat_region
36101..36215
  /note="HALL repeat: matches 27..144 of consensus"
  repeat_region
36283..36679
  /note="match: GSS: Em:AQ679890"
  repeat_region
38239..38719
  /note="match: GSS: Em:AQ241053"
  repeat_region
39043..39393
  /note="L1M4 repeat: matches 5484..5818 of consensus"
  repeat_region
40247..40456
  /note="L1MA5 repeat: matches 6083..6293 of consensus"
  repeat_region
41291..41600
  /note="AluSp repeat: matches 3..311 of consensus"
  repeat_region
41678..41810
  /note="MIR repeat: matches 76..201 of consensus"
  repeat_region
42317..42535
  /note="MLT1G repeat: matches 301..547 of consensus"
  repeat_region
42611..42777
  /note="MLT1G repeat: matches 106..271 of consensus"
  repeat_region
43051..48834
  /note="L1PA5 repeat: matches 361..6143 of consensus"
  repeat_region
48971..49161
  /note="MIR repeat: matches 11..208 of consensus"
  repeat_region
51713..51814
  /note="match: GSS: Em:AQ486704"
  repeat_region
51825..52294
  /note="L1PA7 repeat: matches 5676..6145 of consensus"
  repeat_region
52319..53592
  /note="L1M1 repeat: matches 1384..241 of consensus"
  repeat_region
53593..53889
  /note="AluY repeat: matches 1..298 of consensus"
  repeat_region
53890..54000
  /note="L1M1 repeat: matches 241..126 of consensus"
  repeat_region
54010..54781
  /note="L1PA3 repeat: matches 5376..6146 of consensus"
  repeat_region
54801..54858
  /note="L1M1 repeat: matches 5385..5437 of consensus"
  repeat_region
54871..56799
  /note="L1M1 repeat: matches 130..1815 of consensus"
  repeat_region
56815..57963
  /note="L1PA13 repeat: matches 651..606 of consensus"
  repeat_region
57986..58552
  /note="L1PA13 repeat: matches 758..1326 of consensus"
  repeat_region
58510..58615

```

```
/note="L1 repeat: matches 2444. .2549 of consensus"
58612. .59898
/note="L1MA2 repeat: matches 4985. .6290 of consensus"
60790. .60855
/note="L2 repeat: matches 2644. .2709 of consensus"
60987. .61271
/note="AluX repeat: matches 1. .283 of consensus"
61004. .61457
/note="match: STS: Em:HSC71D4"
61362. .61583
/note="L1MC/D repeat: matches 5422. .5651 of consensus"
61690. .61749
/note="30 copies 2 mer ca 75% conserved"
63509. .63631
/note="AluYa5/8 repeat: matches 187. .309 of consensus"
63758. .63819
/note="31 copies 2 mer ca 83% conserved"
64852. .64887
/note="18 copies 2 mer tt 83% conserved"
64901. .68363
/note="L1PB1 repeat: matches 2691. .6151 of consensus"
68363. .70591
/note="L1PA13 repeat: matches 3895. .6156 of consensus"
70603. .71745
/note="L1M1 repeat: matches 1803. .2941 of consensus"
71727. .73819
/note="L1PBa repeat: matches -279. .2760 of consensus"
73807. .74870
/note="L1PBa repeat: matches -1537. .412 of consensus"
75305. .75479
/note="L1M2 repeat: matches 307. .478 of consensus"
75480. .75832
/note="THEB repeat: matches 1. .364 of consensus"
75833. .76932
/note="L1M2 repeat: matches 478. .1582 of consensus"
76975. .77185
/note="L1M2 repeat: matches 1773. .1587 of consensus"
77293. .77665
/note="MLT-A1 repeat: matches 3. .365 of consensus"
77745. .77964
/note="L1 repeat: matches 1880. .2112 of consensus"
78016. .78085
/note="L1 repeat: matches 2203. .2272 of consensus"
78087. .80520
/note="L1PA5 repeat: matches 3693. .6141 of consensus"
80541. .81238
/note="L1 repeat: matches 2254. .3032 of consensus"
81231. .81559
/note="L1ME1 repeat: matches 5763. .6101 of consensus"
81653. .81904
/note="MIR repeat: matches 7. .262 of consensus"
83147. .83435
/note="AluY repeat: matches 2. .299 of consensus"
83506. .84072
/note="MER7C repeat: matches 38. .602 of consensus"
85050. .85188
/note="L1MA3 repeat: matches 6166. .6304 of consensus"

Query Match      55.9%  Score 954.2; DB 9; Length 200191;
Best Local Similarity 81.9%  Pred. No. 1.4e-210;
Matches 1205; Conservative 7; Mismatches 210; Indels 49; Gaps 8;

QY  98  CCGCTCCGGCGCGCGCATGGAGCTGAGCTGCGGGGTGCGCGGCTCAGCGCGCG 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125618  CCCCACCCCGCGACGCCCATAGAGCTGAGCTGTGGCTTCTCAGGCTCAGCGGTGTG 125559

QY  158  GAGCGCGCGG---ACGTGGATGTGGCGCGCATCTCC-----GCCCTTGCC 201
    ||||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125558  GAGCGCGTGGAAAAGTGGTATGGCAGCATTTCCACCCCGCCCGTCCCGCCACC 125499

QY  202  CCGCGCCCGCGAGCTGGAGCTGTCTCCGGACACAGATATGAAATAGTGTGGAGCTC 261
    ||||| || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125498  CCGCACCCCGAGCTGAGTCTCTTGGACAGATATGAGATGAGTCTTGAAGCTT 125439
```

```
QY  262  GAAGCAATAAAGTTGTTGGGTATCCTGATGATGCGCAATGCTTCAITATTATTATG 321
Db 125438  GAAGCAATAAAGTTGTTGGGTAACTGATGATGCGCAATGCTTCAITATTATTATG 125379

QY  322  AAGTCTCCAAAGCAGTAGCAAGAAAAAATGAAAA----- 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125378  AAGCTTCCAAAGCAGTAGCCAATATAATAATAATAATAATAATAATAATAATAATG 125319

QY  360  -----GGGAAGTAATAATACCCAAAGAGAGTTCTTGGAGATATATACCCCTCCGAGG 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125318  AAGGTGGTTTAATAATAATAATACCCAAAGAGAGCTTTTGGAGATGTTGCTACACCTCTCGAGG 125259

QY  415  CATACTGGAACCGAGACGAGAGAGAGCTGAACCGGAGTACACCCCATCTCTGAGCATGC 474
    || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125258  CACACTGGAACGAGAGACAAAGAAAGCTGAACAAGTGGTACAAATCCCATCTGAGTAGT 125199

QY  475  TGACCAACACGAGCGGGGCGGAGGCTCTCCAATATAAAGCCATCTCAACTACTGCG 534
    || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125198  TGGCCACACGAC- GGGGAAGTGTACGGGTTTTTCAGTATAAGCCATCTCAATTACTGTG 125140

QY  535  AACTGACCTGAGGGTCACCTCGGTGGTTACGGGTTTTAAACAATTCGCGGACAGATTTA 594
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125139  AACTGACTTTCAGGTTTCACGTGAGTGGTTACAGATTTTAAACAATTTCTCTGGACAGATGTG 125080

QY  595  AAGACTTTCTGCTGTTTTCAGATGCGGCAATATTACTGCTTATAGATCAGCCGATA 654
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125079  ATGACTTTCTGATGTTTTCAGATGCTGCAATGATTCACCTGCTTTTAGATCAGCCGATA 125020

QY  655  AGTGTGCAAGAAACCTTCTTGTGCTGGCGATTAAAGTCCCTCACTCCACATTTTGCCA 714
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125019  AGTGTGCAAGAAAGCACITCTTATTTCGTGGTATTAAGTACCTCCTCCCATTTTGCCA 124960

QY  715  GAAGGCAAGCAATCCGGGAATCTCGGGCCCAAGAAAGCAACGAGGAGAACCAACGGTGG 774
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124959  CAAGGCAAGTGATTCGGGATTCCTGGGTGAGAGAAACCAACGTTGGGAACCAAGGGTGA 124900

QY  775  TGGAGCTTCTCGTGGCGGACACACCCCGAGGAGCAACCCGACCTTCAGATA 834
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124899  TGAAGTTTCTTGGTGGTTCAGATACCCCGAGGTCACACCCCGGAGTTTCTGATA 124840

QY  835  TGCTGAAATTTGAGGTGAGAAGCA-CCAAAGACATTTATGTGGAATCAGAGACACT 893
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124839  TGCTGAAATTTGAGATGAGAAGCAACCCAGAGATAGTTTTATGCGGAGCTATAGAGACT 124780

QY  894  TCTTCAACTGTCTCTGAGGAGAGTGTCTTTCTCAGGTGGTAAGTACTTCTCGGCCA 953
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124779  TTCTCAAGTGTCTCTGAAAGAAGTGTGTTTCTC- GCTGGGTAGTACTTCTCGGCCA 124721

QY  954  GACACTGAGTTTGTTCAGGGCGATGACGATGTTTTTGTGAACCCCATCATCCTG 1013
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124720  GACACTGAGTTGTTTTCAGGGCAAGATGATGTTTTTGTGAACCCCATCATCTACTG 124661

QY  1014  AATTACTTGAATAGTTTATCCAAG- ACCAAAGCAAGATCTCTTCATAGGTGATGTAT 1072
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124660  AATTACTGAAATAGCTTATTCAAGAAAAAAGCAAGATCTGTTTATAGTGTATGTAT 124601

QY  1073  CCACATGCTGGACCTCATCGGATAGAGCTGAAGTACTACTATCCAGAACTTGTTA 1132
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124600  CCACATGCTGGACCTCATGGGATAGAGCTGAATACTACTATCTAGAAGTTGTTA 124541

QY  1133  CTCTGGCTCTACCCACCTTATGCA- GGGGAGGGGGTTCCTCTACTCCGCCACCTG 1190
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124540  CTCTGTCTCTACTCACCTTATGCAAGGGGGGGGGATTCCTCTAGTCCAGCCACCTG 124481

QY  1191  GCCTGAGGCTGATACCATATCACTGACAGGTCCATCTCTACCCCATTTGATGACGTTAT 1250
    || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124480  GCTCTGAGGCTGATACCATATCACTGACTGGTCCATCTCTAACTTATGATGATTTTAT 124421

QY  1251  ACTGGAATGTGCCTTCAGAAACTCGGCCTCGTTCCAGAGAAACACAAAGGCTTCAGGACA 1310
    ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124420  ACTGCAATGTGCCTTCAAAAACCTAGCTGTTTCCAGAGAAGACAAAGGCTTCAGGACA 124361
```



```
* 160986 161087: gap of 100 bp
* 161088 164251: contig of 3684 bp in length
* 164752 164851: gap of 100 bp
* 164852 167410: contig of 2559 bp in length
* 167411 167510: gap of 100 bp
* 167511 171050: contig of 3540 bp in length
* 171051 171150: gap of 100 bp
* 171151 173702: contig of 2552 bp in length
* 173703 173802: gap of 100 bp
* 173803 175996: contig of 2194 bp in length
* 175997 176096: gap of 100 bp
* 176097 180350: contig of 4254 bp in length
* 180351 180450: gap of 100 bp
* 180451 183955: contig of 3505 bp in length
* 183956 184055: gap of 100 bp
* 184056 191665: contig of 7610 bp in length
* 191666 191765: gap of 100 bp
* 191766 200661: contig of 8896 bp in length
* 200662 200761: gap of 100 bp
* 200762 203552: contig of 2791 bp in length
* 203553 203652: gap of 100 bp
* 203653 205703: contig of 2051 bp in length
* 205704 205804: gap of 100 bp
* 205804 208860: contig of 3057 bp in length
* 208861 208960: gap of 100 bp
* 208961 215264: contig of 6304 bp in length
* 215265 215364: gap of 100 bp
* 215365 225728: contig of 10364 bp in length
* 225729 225828: gap of 100 bp
* 225829 227836: contig of 2008 bp in length
* 227837 227936: gap of 100 bp
* 227937 231580: contig of 3644 bp in length
* 231581 231680: gap of 100 bp
* 231681 235368: contig of 3688 bp in length
* 235369 235468: gap of 100 bp
* 235469 238205: contig of 2737 bp in length
* 238206 238305: gap of 100 bp
* 238306 240714: contig of 2409 bp in length
* 240715 240814: gap of 100 bp
* 240815 250346: contig of 9432 bp in length
* 250347 250346: gap of 100 bp
* 250347 253276: contig of 4830 bp in length
* 253277 255276: gap of 100 bp
* 255277 258769: contig of 3493 bp in length
* 258770 258869: gap of 100 bp
* 258870 261419: contig of 2550 bp in length
* 261420 261519: gap of 100 bp
* 261520 265240: contig of 3721 bp in length
* 265241 265341: gap of 100 bp
* 265341 267630: contig of 2290 bp in length
* 267631 267730: gap of 100 bp
* 267731 269846: contig of 2116 bp in length
* 269847 269946: gap of 100 bp
* 269947 274498: contig of 4552 bp in length
* 274499 274598: gap of 100 bp
* 274599 277625: contig of 3027 bp in length
* 277626 277725: gap of 100 bp
* 277726 280792: contig of 3057 bp in length
* 280793 280892: gap of 100 bp
* 280893 285583: contig of 4691 bp in length
* 285584 285684: gap of 100 bp
* 285684 288669: contig of 2986 bp in length.
FEATURES
    source
        1..288669
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosomes="10"
            /clone="Rp13-1M23"
            /clone_lib="RPCI-13.1"
            1..5254
            /note="assembly_fragment:01614
            fragment_chain:1"
misc_feature
    1014 AATTACTTGAATAGTTTATCCAAAG-ACCAAGGCCAAGAGTCTCTTCAAGTGTGAT
    1072
    251898 AATTACGTGAATAGTCTTATTCAGAAAAAAGCCAAAGATCTGTTTATAGTGTGAT 251839
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 12:46:17 ; Search time 482.17 Seconds
(without alignments)
9556.669 Million cell updates/sec

Title: US-09-804-357b-13

Perfect score: 1707
Sequence: 1 acgcgtccgcagcggcag.....aatctgttgatggccctt 1707

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1694.8	99.3	1707	AA12716	CDNA encoding huma
2	1694.8	99.3	1707	AA167869	Human LIG46 poly
3	1694.8	99.3	1707	AA167869	Human LIG46 cDNA.
4	1531.4	89.7	1831	AA167869	Human beta3Gnt gen
5	1504.2	88.1	1912	AA167869	Human beta 1,3-N-a
6	1488	87.2	2745	AA167869	Human cardiac and
7	1231.4	72.1	2442	AA167869	Human cDNA sequenc
8	1229.8	72.0	2536	AA167869	Human Brainiac cDN

9	1036.4	60.8	1194	21	AA288487	Human brainiac pro
10	1002.2	58.7	2676	21	AA288486	Murine brainiac en
11	1002.2	58.7	2676	21	AA287185	Murine Brainiac cD
12	931.4	54.6	1260	25	ABV75081	Murine beta3Gnt ge
13	906.4	53.1	1196	21	AA12714	CDNA encoding mur
14	906.4	53.1	1196	23	AA167865	Murine LIG46 poly
15	906.4	53.1	1196	23	AA167866	Murine LIG46 cDNA.
16	906.4	53.1	1194	23	AA167866	Murine LIG46 poly
17	717.2	42.0	932	22	AAH07800	Human cDNA clone (
18	533.6	31.3	556	20	AAH89848	EST clone CW1306.
19	431.4	25.3	857	19	AAV66384	CDNA clone W26453.
20	328.2	19.2	480	19	AAV66390	CDNA clone H47991.
21	284.8	16.7	303	19	AAV66395	CDNA clone H00589.
22	279	16.3	282	19	AAV66398	CDNA clone H00116.
23	279	16.3	282	19	AAV66375	CDNA clone HTAAW41
24	267.4	15.7	428	19	AAV66400	CDNA clone AA38183
25	264	15.5	266	19	AAV66401	CDNA clone AA37708
26	264	15.5	266	19	AAV66376	CDNA clone HTABE60
27	257	15.1	259	19	AAV66402	CDNA clone H13126
28	257	15.1	259	19	AAV66378	CDNA clone H518A68
29	223.4	13.1	473	21	AA288389	Human secreted pro
30	203.4	11.9	1245	24	ABK89794	Human galactosyltr
31	175.8	10.3	368	19	AAV66404	CDNA clone AA37708
32	174.4	10.2	1134	21	AA58791	DNA encoding the b
33	174.4	10.2	1134	24	ABS53140	Degenerate DNA enc
34	174.4	10.2	1134	24	ABD39776	Human zsssp6 degen
35	167.8	9.8	1360	24	AA235225	Human TRNFR-10 cDN
36	167.6	9.8	1336	24	ABX97158	Human NOV106a cDNA
37	167.6	9.8	1359	23	AA584438	DNA encoding novel
38	166.2	9.7	3356	25	ABX63848	Human cDNA #848 d1
39	164.8	9.7	2003	24	ABK89793	CDNA encoding huma
40	163.2	9.6	1116	19	AAV49599	Human epidermold c
41	163.2	9.6	1208	24	AA516945	Human beta1,3-N-ac
42	163.2	9.6	2186	24	ABL49598	Human epidermold c
43	163.2	9.6	2186	24	ABL41994	Nucleotide sequenc
44	163.2	9.6	2186	25	ABZ78132	Human cancer-relat
45	163.2	9.6	2186	25	ABV76854	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AA12716
ID AA12716 standard; cDNA; 1707 BP.
AC AA12716;
XX 25-JUL-2000 (first entry)
XX CDNA encoding human LIG46 polypeptide.
XX Leptin: LIG46; body weight; leptin inducible gene: obesity; cachexia;
XX LIG56; tgtp; LRG-47; KC10-II; Stral3; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 245..1439
XX /*tag= a
XX /product= *LIG46*
XX WO200015826-A2.
XX PD 23-MAR-2000.
XX PF 10-SEP-1999; 99WO-US20722.
XX PR 10-SEP-1998; 98US-0150857.
XX PR 29-OCT-1998; 98US-0106378.
XX PR 19-NOV-1998; 98US-0195896.
XX PR 15-APR-1999; 99US-0292228.
XX

Db 1621 NTTATGGTGAATGGCAGGATATTGGTCTGACTTACCGNTAGGGGANTTTAAACITGGNC 1680

Qy 1681 CTTTTGAATCTGTTGGATGGCCCTT 1707

Db 1681 CTTTTGAATCTGTTGGATGGCCCTT 1707

RESULT 2

AA167869

ID AA167869 standard: cDNA: 1707 BP.

XX AA167869;

DT 13-MAR-2002 (first entry)

XX Human LIG46 polypeptide encoding cDNA.

XX Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Stral3;
KW anorectic; anabolic; antisense therapy; human; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 246..1439

FT /*tag= a

FT /product= "LIG46"

XX US2001024808-A1.

XX 27-SEP-2001.

XX 12-MAR-2001; 2001US-0804357.

XX 29-OCT-1998; 98US-106378P.

PR 19-NOV-1998; 98US-019589S.

PR 10-SEP-1998; 98US-0150857.

XX (MILL-) MILLENNIUM PHARM INC.

XX White D, Zhou J, Tartaglia LA;

XX WPI; 2001-624963/72.

DR P-PSDB; AAG66118.

XX An isolated polypeptide useful for treating a weight disorder and for
PT screening compounds that may inhibit protein activity comprises a
PT leptin-induced protein -

XX Example 2; Fig 7; 46pp; English.

XX The invention relates to genes whose expression are induced by leptin.
CC LIG46 and LIG56 are novel leptin induced genes (Lig), while four other
CC genes such as Tgtp, LRG-47, RC10-II and Stral3 have been previously
CC identified. The leptin induced proteins can be expressed by standard
CC recombinant methodology. The proteins and encoding polynucleotides may
CC be used in screening assays to identify compounds that may bind to it.
CC Administering a molecule, e.g., an antisense molecule, which reduces
CC expression of activity of protein selected from LIG46, LIG56, Tgtp,
CC LRG-47, RC10-II, and Stral3 may be used to treat a weight disorder. The
CC method may further comprise administering leptin. The present sequence
CC represents a cDNA encoding a human LIG46 protein.

XX Sequence 1707 BP; 452 A; 403 C; 435 G; 400 T; 17 other:

Query Match 99.3%; Score 1694.8; DB 23; Length 1707;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACGGGTCCGGCAGCGGCGGAGCAGCGGCAACAAGTCCGGAGGCTAGCAGAGCCA 60
|||||

Db 1 ACGGTCCGGCAGCGGCGGAGCAGCGGCAACAAGTCCGGAGGCTAGCAGAGCCA 60
|||||

Qy 61 ACGGCGACAGTCCCTGCGCGGACACCGCGGCGCCCGCTCCGGGCGCGGCATGG 120

Db 61 AGCCGAGCAGTCCCTGCGCGCAGACACCGCGCGCGCGTCCGGGCGCGGCATGG 120
|||||

Qy 121 AGCGTGAGTGGCGGGTGGCGGGTGGAGCCCGCGGAGCGCGCGGGAGTGGATGG 180
|||||

Db 121 AGCGTGAGTGGCGGGTGGCGGGTGGAGCCCGCGGAGCGCGCGGGAGTGGATGG 180
|||||

Qy 181 CCGCGATCTCCCGCCCTTGGCCCGCCCGCGAGCTGGAGCTGCTCCCGACAAGATAT 240
|||||

Db 181 CCGCGATCTCCCGCCCTTGGCCCGCCCGCGAGCTGGAGCTGCTCCCGACAAGATAT 240
|||||

Qy 241 GAGAAATGAGTGTGGACGTCGAAAGATAAAGTTGTTGGGTATCTCTGATGGCAATG 300
|||||

Db 241 GAGAAATGAGTGTGGACGTCGAAAGATAAAGTTGTTGGGTATCTCTGATGGCAATG 300
|||||

Qy 301 TCTTCATTTATTTATTTATGGAAGTCTCCAAAAGCAGTAGCCCAAGAAAAAATGGAAG 360
|||||

Db 301 TCTTCATTTATTTATTTATGGAAGTCTCCAAAAGCAGTAGCCCAAGAAAAAATGGAAG 360
|||||

Qy 361 GGAAGTATAATACCCARAGAGAGTCTTGGAAGATATCTACCCCTCCCGAGGCATCT 420
|||||

Db 361 GGAAGTATAATACCCARAGAGAGTCTTGGAAGATATCTACCCCTCCCGAGGCATCT 420
|||||

Qy 421 GGAACCGAGAGCAAGAGAGCTGAACCGGCGAGTACAACCCCATCTCTGAGCATCTGACCA 480
|||||

Db 421 GGAACCGAGAGCAAGAGAGCTGAACCGGCGAGTACAACCCCATCTCTGAGCATCTGACCA 480
|||||

Qy 481 ACCAGACGGGGAGGGGCGGAGGCTCTCCAATATAAGCCATCTGAACCTAGTGGAACTG 540
|||||

Db 481 ACCAGACGGGGAGGGGCGGAGGCTCTCCAATATAAGCCATCTGAACCTAGTGGAACTG 540
|||||

Qy 541 ACTGAGGGTCACGTCGGTGTACGGGTTTAACTTTTAACTTTGCGGACAGATTTAAAGACT 600
|||||

Db 541 ACTGAGGGTCACGTCGGTGTACGGGTTTAACTTTTAACTTTGCGGACAGATTTAAAGACT 600
|||||

Qy 601 TTCTGCTGTTATTTGAGATGCGCAATTAATCACTGCTTATAGATCAGCCGGATGAAGTGTG 660
|||||

Db 601 TTCTGCTGTTATTTGAGATGCGCAATTAATCACTGCTTATAGATCAGCCGGATGAAGTGTG 660
|||||

Qy 661 CAAAGAAACCTTTCTTCTGCTGGCGATTAACTTAACTTCTCCACATTTTGCCAGAAGGC 720
|||||

Db 661 CAAAGAAACCTTTCTTCTGCTGGCGATTAACTTAACTTCTCCACATTTTGCCAGAAGGC 720
|||||

Qy 721 AGCAATCCGGGATCTCTGGGCGCAAGAAACACGAGGAGCAACCGTGGTGGAG 780
|||||

Db 721 AGCAATCCGGGATCTCTGGGCGCAAGAAACACGAGGAGCAACCGTGGTGGAG 780
|||||

Qy 781 TCTTCTGCTGGCGACACACCCCGAGGACACACCCCGACCTTTTCAGATATGCTGA 840
|||||

Db 781 TCTTCTGCTGGCGACACACCCCGAGGACACACCCCGACCTTTTCAGATATGCTGA 840
|||||

Qy 841 AATTTGAGATGAGAACCAAGACATCTTATGTGGAACCTACAGAGACACTTTCTTCA 900
|||||

Db 841 AATTTGAGATGAGAACCAAGACATCTTATGTGGAACCTACAGAGACACTTTCTTCA 900
|||||

Qy 901 ACTTGCTCTGAGGAGAGTCTGTTCTTCAGGTGGGTAACTACTCTCTGCCAGACACTG 960
|||||

Db 901 ACTTGCTCTGAGGAGAGTCTGTTCTTCAGGTGGGTAACTACTCTCTGCCAGACACTG 960
|||||

Qy 961 AGTTTGTGTTTCAAGGGCGATGACGATGTTTGTGTGAACCCCATCAGATCTCTGAAITACT 1020
|||||

Db 961 AGTTTGTGTTTCAAGGGCGATGACGATGTTTGTGTGAACCCCATCAGATCTCTGAAITACT 1020
|||||

Qy 1021 TGAATAGTTTATCCAGACCAAGCCAAAGATCTCTTCATAGGTGATGATCCCAATG 1080
|||||

Db 1021 TGAATAGTTTATCCAGACCAAGCCAAAGATCTCTTCATAGGTGATGATCCCAATG 1080
|||||

Qy 1081 CTGGACCTCATCGGGATGAAGCTGAAGTACTACATCTCCAGAGAGTGTGTTACTCTGGCC 1140
|||||

Db 1081 CTGGACCTCATCGGGATGAAGCTGAAGTACTACATCTCCAGAGAGTGTGTTACTCTGGCC 1140
|||||

Qy 1141 TCTACCCACCTCATGACGGGGAGGGGTTCTCTACTCCGCCACCTGGCCCTGAGGC 1200
|||||

Db 1141 TCTACCCACCTATGACGGGGAGGGGGTTCCTCTACTCGGCCACCTGGCCCTGAGGC 1200
 Qy 1201 TGACCATATCACTGACCGAGTCCATCTCTACCCCATGTAIGACGTTATATAGTAATGT 1260
 Db 1201 TGTACCATATCACTGACCGAGTCCATCTCTACCCCATGTAIGACGTTATATAGTAATGT 1260
 Qy 1261 GCCTTCAGAACTCGGCTCGTTCAGAGAAACACAAAGGCTTCAGGACATTTGATATCG 1320
 Db 1261 GCCTTCAGAACTCGGCTCGTTCAGAGAAACACAAAGGCTTCAGGACATTTGATATCG 1320
 Qy 1321 AGAGAAAAACAATAAATACATCTGCTCTCTATAGATCTGATGTTAGTACATAGTAA 1380
 Db 1321 AGAGAAAAACAATAAATACATCTGCTCTCTATAGATCTGATGTTAGTACATAGTAA 1380
 Qy 1381 AACCTCAAGAGATGATGATATTTGCTCAGTTGCAGAGTCTCATTTAAATGCTAAA 1440
 Db 1381 AACCTCAAGAGATGATGATATTTGCTCAGTTGCAGAGTCTCATTTAAATGCTAAA 1440
 Qy 1441 ATAGATACAACTCAATTTKGSATWGRAAGGGGTWTTTGRATWGYCCCATGTTGGGT 1500
 Db 1441 ATAGATACAACTCAATTTKGSATWGRAAGGGGTWTTTGRATWGYCCCATGTTGGGT 1500
 Qy 1501 CTCACATAGAGTAATTTCTATTNAANCATGAATGGCTTTATGAGTATACCATTT 1560
 Db 1501 CTCACATAGAGTAATTTCTATTNAANCATGAATGGCTTTATGAGTATACCATTT 1560
 Qy 1561 ANGCGCTCTAANCCTTCATTGTCAGTGAAGAGGAAAGCGGAGAGTAATTT 1620
 Db 1561 ANGCGCTCTAANCCTTCATTGTCAGTGAAGAGGAAAGCGGAGAGTAATTT 1620
 Qy 1621 NTTTTGGTGAATGGCAGGATATGGTCTGACTTACCGNTAGGGGANTTTAAACTGGNC 1680
 Db 1621 NTTTTGGTGAATGGCAGGATATGGTCTGACTTACCGNTAGGGGANTTTAAACTGGNC 1680
 Qy 1681 CTTTTTGAATCTGTTTGGATGGCCCTT 1707
 Db 1681 CTTTTTGAATCTGTTTGGATGGCCCTT 1707
 RESULT 3
 AAD46673
 ID AAD46673 standard; cdna; 1707 BP.
 AC AAD46673;
 XX 27-JAN-2003 (first entry)
 DT Human LIG46 cdna.
 DE
 XX LIG46; leptin; low body weight; chromosomal mapping; tissue typing;
 KW forensic biology; transgenic; gene therapy; antianorectic; human;
 KW chromosome 2; gene; ss.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT CDS 246..1439
 FT /*tag= a
 FT /product= "Human LIG46 protein"
 FT sig_peptide 246..341
 FT /*tag= b
 FT mat_peptide 342..1436
 FT /*tag= c
 FT /product= "Human mature LIG46 protein"
 FT
 XX WO200274905-A2.
 PN
 XX 26-SEP-2002.
 PD
 XX 20-NOV-2001; 2001WO-US43345.
 PF
 XX 21-NOV-2000; 2000US-0717778.
 PR
 XX

PA (MILL-) MILLENNIUM PHARM INC.
 XX White DW, Zhou J, Tartaglia LA, Stricker-krongrad A, Clausen H;
 P2 WPI; 2002-759886/82.
 DR P-PSDB: AAE29093.
 XX New isolated nucleic acid molecules, designated as leptin-induced genes
 PT 45 (LIG46), useful for treating a subject having a disorder
 PT characterized by undesirable level of LIG46 expression or activity,
 PT such as low body weight
 XX
 PS Claim 15; Fig 5; 90pp; English.
 XX The invention relates to LIG46, a gene whose expression is induced by
 CC leptin. LIG46 DNA and protein are useful in treating a subject having
 CC a disorder characterized by undesirable level of LIG46 expression or
 CC activity, such as low body weight. They are also useful in a screening
 CC assay, chromosomal mapping, tissue typing and forensic biology. The
 CC probes based on the LIG46 nucleotide sequence are useful for detecting
 CC transcripts or genomic sequences encoding the same or related proteins.
 CC The LIG46 polypeptides are useful as immunogens for raising anti-LIG46
 CC antibodies. The host cells are useful for producing non-human transgenic
 CC animals. LIG46 DNA is used in gene therapy. The present sequence is
 CC human LIG46 cdna. LIG46 gene is located at chromosome 2.
 XX
 SQ Sequence 1707 BP; 452 A; 403 C; 435 G; 400 T; 17 other;
 Query Match 99.3%; Score 1694.8; DB 24; Length 1707;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ACGGGTCCGCGCAGCGGCGAGCGGCGGCAACAAAGTCCCGAGGCTAGCAGACCCA 60
 Db 1 ACGGGTCCGCGCAGCGGCGAGCGGCGGCAACAAAGTCCCGAGGCTAGCAGACCCA 60
 Qy 61 AGCGGAGCAGTCCCTCCGCGCAGCACCGCGGCGCGCGCTCCGGGGCGCGGCATGG 120
 Db 61 AGCGGAGCAGTCCCTCCGCGCAGCACCGCGGCGCGCGCTCCGGGGCGCGGCATGG 120
 Qy 121 AGCGTGAGCTCGGCGGTCGCGGCGTGCAGCGCGGCGGAGCGCGGAGCTGATCTGG 180
 Db 121 AGCGTGAGCTCGGCGGTCGCGGCGTGCAGCGCGGCGGAGCGCGGAGCTGATCTGG 180
 Qy 181 CGCGCATCTCCCGCCCTTGCCCGCGCGCGCGCGCGAGCTGGAGCTGCTCCCGACAAATAT 240
 Db 181 CGCGCATCTCCCGCCCTTGCCCGCGCGCGCGCGAGCTGGAGCTGCTCCCGACAAATAT 240
 Qy 241 GAGAAATGAGTGTGGAGCTCGAAGATTAAGTTGTTGGGTATCTCTGATGTGCAATG 300
 Db 241 GAGAAATGAGTGTGGAGCTCGAAGATTAAGTTGTTGGGTATCTCTGATGTGCAATG 300
 Qy 301 TCTTCATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 360
 Db 301 TCTTCATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 360
 Qy 361 GGGAAAGTAAATATCCCAAGAGAGTCTCGAAGAGTCTCCAAAGCAGTAGCCAAAGAAATG 420
 Db 361 GGGAAAGTAAATATCCCAAGAGAGTCTCGAAGAGTCTCGAAGAGTCTCCAAAGCAGTAGCC 420
 Qy 421 GGAACCGAGAGCAAGAGAGCTGACCGGCGAGTACACCCCATCTCTGAGCATCTGACCA 480
 Db 421 GGAACCGAGAGCAAGAGAGCTGACCGGCGAGTACACCCCATCTCTGAGCATCTGACCA 480
 Qy 481 ACCAGACGGGAGGCGGCGAGGCTCTCCAATATAAGCCATCTGAACTACTGCGAACCTG 540
 Db 481 ACCAGACGGGAGGCGGCGAGGCTCTCCAATATAAGCCATCTGAACTACTGCGAACCTG 540
 Qy 541 ACCTGAGGTCACGTGCGTGGTTACGGGTTTTAAACAATTTGCCGCGACAGATTTAAAGACT 600
 Db 541 ACCTGAGGTCACGTGCGTGGTTACGGGTTTTAAACAATTTGCCGCGACAGATTTAAAGACT 600
 Qy 601 TTCTGCTGATTTGAGATGCCGCAATTTATTCACCTGCTTATAGATCAGCCGGATAAGTGTG 660


```
QY 133 GCGGTGCGGGCTGAGCGCGGAGCGCGGGAGCTGGATGCTGCGCGGATCTCCC 192
DB 133 GCGGTGCGGGCTGAGCGCGGAGCGCGGGAGCTGGATGCTGCGCGGATCTCCC 182
QY 193 GCCTTGGCCCCGCCGCCGAGCTGAGCTGCTCCCGGACAAGATATGAGAAATGAGTG 252
DB 183 GCCTTGGCCCCGCCGCCGAGCTGAGCTGCTCCCGGACAAGATATGAGAAATGAGTG 242
QY 253 TTGGAGCTGGAAGATAAAGTTGTTGGGTATCCTGATGCGCAATGCTTCTTATT 312
DB 243 TTGGAGCTGGAAGATAAAGTTGTTGGGTATCCTGATGCGCAATGCTTCTTATT 302
QY 313 TTATTATGGAAGTCTCCAAAGCAGTAGCCAAAGAAAATGGAAGGGGAAGTAATAA 372
DB 303 TTATTATGGAAGTCTCCAAAGCAGTAGCCAAAGAAAATGGAAGGGGAAGTAATAA 362
QY 373 TACCCAAAGAGAAAGTTCTGGAGATATCTACCCCTCCCGAGGCATCTGGAACCGAGGC 432
DB 363 TACCCAAAGAGAAAGTTCTGGAGATATCTACCCCTCCCGAGGCATCTGGAACCGAGGC 422
QY 433 AGAGAAAGCTGAACGGGAGTACACCCCATCCTGAGCATGCTGACCAACGAGCGGG 492
DB 423 AGAGAAAGCTGAACGGGAGTACACCCCATCCTGAGCATGCTGACCAACGAGCGGG 482
QY 493 AGCGGGGAGGCTCTCCAAATATAAGCCATCTGAACCTACTGCGAAGCTGAGGGTCA 552
DB 483 AGCGGGGAGGCTCTCCAAATATAAGCCATCTGAACCTACTGCGAAGCTGAGGGTCA 542
QY 553 CGTCGGTGGTTACGGGTTTAAACACTTGGCGGAGAGATTTAAAGACTTTCTGCTGTAIT 612
DB 543 CGTCGGTGGTTACGGGTTTAAACACTTGGCGGAGAGATTTAAAGACTTTCTGCTGTAIT 602
QY 613 TGAGATGCGCAATTTCACTGCTTATAGATCAGCGGATAGTGTGCAAGAAACCT 672
DB 603 TGAGATGCGCAATTTCACTGCTTATAGATCAGCGGATAGTGTGCAAGAAACCT 662
QY 673 TCTTGTGCTGCGATTAAGTCCCTCACTCCACATTTTCCAGAGGCAAGCAATCCGGG 732
DB 663 TCTTGTGCTGCGATTAAGTCCCTCACTCCACATTTTCCAGAGGCAAGCAATCCGGG 722
QY 733 ATCTCGGGCCAAAGAAACAGGAGGAGCAACCAACCGTGGTGGAGTCTTCTGCTGG 792
DB 723 ATCTCGGGCCAAAGAAACAGGAGGAGCAACCAACCGTGGTGGAGTCTTCTGCTGG 782
QY 793 GCCAGACACCCCGAGGAGCAACACCCCGACCTTTCAGATATGCTGAAATTTGAGAGTG 852
DB 783 GCCAGACACCCCGAGGAGCAACACCCCGACCTTTCAGATATGCTGAAATTTGAGAGTG 842
QY 853 AGAAGCACCAAGACATTTCTATGTGGAATACAGAGACACTTTCTTCAACTTGTCTGTA 912
DB 843 AGAAGCACCAAGACATTTCTATGTGGAATACAGAGACACTTTCTTCAACTTGTCTGTA 902
QY 913 AGAAGTCTGTTCTCAGGTGGTAAAGTACTTCTGCCAGACACTGAGTTGTCTCA 972
DB 903 AGAAGTCTGTTCTCAGGTGGTAAAGTACTTCTGCCAGACACTGAGTTGTCTCA 962
QY 973 AGGGGATGACATGTTTTTGTGAACACCCATCATCTCTGAAATTTGAAATGTTAT 1032
DB 963 AGGGGATGACATGTTTTTGTGAACACCCATCATCTCTGAAATTTGAAATGTTAT 1022
QY 1033 CCAAGACCAACCAAGATCTCTTATAGTGTGATGCCACATGCTGAGCTCATCA 1092
DB 1023 CCAAGACCAACCAAGATCTCTTATAGTGTGATGCCACATGCTGAGCTCATCA 1082
QY 1093 GGGATAAGAGCTGAAGTACTACATCCAGAGAGTTGTTTACTCGGCTCTACCAACCT 1152
DB 1083 GGGATAAGAGCTGAAGTACTACATCCAGAGAGTTGTTTACTCGGCTCTACCAACCT 1142
QY 1153 ATCGAGGGGAGGGGTTCTCTACTCGGGCCACTGCGCTGAGGCTGACCATATCA 1212
DB 1143 ATCGAGGGGAGGGGTTCTCTACTCGGGCCACTGCGCTGAGGCTGACCATATCA 1202
```

```
QY 1213 CTGACAGGTCCTCTACCCCATTTGATCAGCTTTTATCTACTGGAATGTGCTTCAGAAAC 1272
DB 1203 CTGACAGGTCCTCTCTACCCCATTTGATCAGCTTTTATCTACTGGAATGTGCTTCAGAAAC 1262
QY 1273 TCGGCTCGTTCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGGAGAAAAACA 1332
DB 1263 TCGGCTCGTTCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGGAGAAAAACA 1322
QY 1333 AAAATACATCTGCTCTATGTAGATCTGATGTAGTACATAGTACAGAAACCTCAGAGA 1392
DB 1323 AAAATACATCTGCTCTATGTAGATCTGATGTAGTACATAGTACAGAAACCTCAGAGA 1382
QY 1393 TGATTCATATTTGGTCTCAGTTCAGAGTCTCAITTTAAATGCTAAAATAGATACAAAC 1452
DB 1383 TGATTCATATTTGGTCTCAGTTCAGAGTCTCAITTTAAATGCTAAAATAGATACAAAC 1442
QY 1453 TCAATTTKSATWGRAGGGGTTTGTGATWGGYCCCATGTTGGGCTCTCACAATAGAG 1512
DB 1443 TCAATTTGCTAGAAAGGTGATTTTGAATAGTTCCTCATGTTGTTCTCATTAGAG 1502
QY 1513 TAATTTCTATTTNAAC--ATGAAATTTGCTTTTATGAGTGATACCCATTT-ANGGCTCT 1569
DB 1503 TAATTTCTATTTAAACCATGAAATTTGCTTTTATGAGTGATACCCATTTGAGGGCTCT 1562
QY 1570 AANCCCTT--CATTTGNACTCAGCTGAAGGAGGAAAGCGGAGAGGTAATTTTAT 1626
DB 1563 AANCCCTTCAATTTGGTACTCAGCTGAAG-AGGGAAGCGGAGATGTTAATTTTTTT 1621
QY 1627 GGTGA 1631
DB 1622 ATGGA 1626

RESULT 5
AAF29255
ID AAF29255 standard; cDNA; 1912 BP.
XX
AC AAF29255;
XX
XX 17-APR-2001 (first entry)
XX
DE Human beta 1,3-N-acetylglucosamine transferase G3 cDNA sequence.
XX
XX Sugar chain synthesis agent; beta 1,3-N-acetylglucosamine transferase;
XX inflammation; cancer; metastasis; human; ss.
XX
OS Homo sapiens.
XX
XX WO200100848-A1.
XX
XX 04-JAN-2001.
XX
XX 29-JUN-2000; 2000WO-JP04304.
XX
XX 29-JUN-1999; 99JP-0183437.
XX
XX 16-MAR-2000; 2000JP-0074757.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX Sasaki K, Shiraishi N, Natsume A, Yamada Y, Nakagawa S, Sekine S;
XX
XX WPI; 2001-102895/11.
XX
XX P-PSDB: AAB49748.
XX
PT New polypeptide having beta1,3-N-acetylglucosamine transferase activity
PT for diagnosis of inflammation, cancer and cancer metastasis,
PT development of remedies, and for producing glycoconjugates.
XX
XX Example 2; Page 159-164; 195pp; Japanese.
XX
XX This invention relates to a sugar chain synthesizing agent that contains
XX a polypeptide as the active ingredient, where the polypeptide has beta
XX 1,3-N-acetylglucosamine transferase activity. The polypeptides, DNA
```



```
Db 508 AGCAACGACGAGGAAACCAACCGTGGTGGAGCTCTCTGCTGGCCAGACACCCCCAGAG 567
Qy 810 GACAACACCCCGACCTTTGAGATATCTGTAATTTGAGAGTGAAGACCAAGACATT 859
Db 568 GACAACACCCCGACCTTTGAGATATCTGTAATTTGAGAGTGAAGACCAAGACATT 627
Qy 870 CTTATGTGGAACTACAGACACTTTCTTCAACTTGTCTGGAAGGAAGTCTGTTTC 929
Db 528 CTTATGTGGAACTACAGACACTTTCTTCAACTTGTCTGGAAGGAAGTCTGTTTC 687
Qy 930 AGTGGTAACTACTTCTGCCACACACTGAGTTCTTTCAAGGGCGATGACGATGT 989
Db 688 AGTGGTAACTACTTCTGCCACACACTGAGTTCTTTCAAGGGCGATGACGATGT 747
Qy 990 TTTGTGAACCCCATCATCTCTGAATTTACTTGAATAGTTTATCCAAAGACCAAGCCAAA 1049
Db 748 TTTGTGAACCCCATCATCTCTGAATTTACTTGAATAGTTTATCCAAAGACCAAGCCAAA 807
Qy 1050 GATCTCTTATAGTGTGATGCCACAATGCTGGACCTCATCGGGATAGAAGCTGAAG 1109
Db 808 GATCTCTTATAGTGTGATGCCACAATGCTGGACCTCATCGGGATAGAAGCTGAAG 867
Qy 1110 TACTACATCCAGAAAGTTGTTTACTCTGGCTCTACCCACCTATGCGAGGGGAGGGGG 1169
Db 868 TACTACATCCAGAAAGTTGTTTACTCTGGCTCTACCCACCTATGCGAGGGGAGGGGG 927
Qy 1170 TTCTCTACTCGGCCACCTGGCCCTGAGCTGTACCATATCACTGACGAGGTCATCTC 1229
Db 928 TTCTCTACTCGGCCACCTGGCCCTGAGCTGTACCATATCACTGACGAGGTCATCTC 987
Qy 1230 TACCCCATTTGATGACGTTTATCTGGAATGTGCTTCAGAAACTCGGCCCTGTTCCAGAG 1289
Db 988 TACCCCATTTGATGACGTTTATCTGGAATGTGCTTCAGAAACTCGGCCCTGTTCCAGAG 1047
Qy 1290 AAACACAAGCCTTCAGACATTTGATATGATGAGAGAGAAACAAATAATACATCTGCTC 1349
Db 1048 AAACACAAGCCTTCAGACATTTGATATGATGAGAGAGAAACAAATAATACATCTGCTC 1107
Qy 1350 TATGTAGATCTGATGTTAGTACATAGTAGAAACCTCAAGAGATGATTGATTTGGTCT 1409
Db 1108 TATGTAGATCTGATGTTAGTACATAGTAGAAACCTCAAGAGATGATTGATTTGGTCT 1167
Qy 1410 CAGTTGACAGTCTCATTTAAATGCTAAATAGATACAAACCTCAATTTKGSATWGRAA 1469
Db 1168 CAGTTGACAGTCTCATTTAAATGCTAAATAGATACAAACCTCAATTTGTCATAGAA 1227
Qy 1470 GGGGTWTTTGTATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1529
Db 1228 GGGGTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1287
Qy 1530 --ATGAAATTCCTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1583
Db 1288 CATGAAATTCCTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1347
Qy 1584 ACTCAGCTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1531
Db 1348 ACTCAGCTGAAG-AGGGAAGCGGAGATGATGATGATGATGATGATGATGATGATGATGAT 1394
```

RESULT 9

AAZ88487

ID AAZ88487 standard; cDNA; 1194 BP.

XX

AC

XX

DT

12-MAY-2000 (first entry)

XX

Human brainiac protein encoding cDNA sequence.

DE

KW

Brainiac: egghad; cell adhesion; cytosolic; dermatological;

KW

neuroactive; cell viability; cell growth regulation; psoriasis; cancer;

XX

cell fate specification; skin lesion; nervous system defect; ss.

QS Homo sapiens.

XX CA2225126-AL.

PN XX

XX 17-JUN-1999.

PD XX

XX 17-DEC-1997; 97CA-2225126.

PF XX

XX 17-DEC-1997; 97CA-2225126.

PR XX

XX {HSCR-} HSC RES & DEV LP.

PA Egan SE;

PI WPI: 2000-161481/15.

XX P-PSDB; AAY79954.

DR Mammalian EGGHEAD and BRAINAC proteins which mediate cell to cell

XX lesions and nervous system defects or diseases -

XX Example 2; Page 27; 30pp; English.

PS The present invention describes mammalian EGGHEAD and BRAINAC proteins,

XX which mediate cell-to-cell adhesion. The mammalian proteins are similar

CC to the EGGHEAD and BRAINAC genes of Drosophila melanogaster. BRAINAC

CC and EGGHEAD genes regulate adhesion between epithelial cells and require

CC the presence of the Notch protein for this activity. Mutational

CC disruption of BRAINAC, EGGHEAD or Notch results in the loss of follicle

CC epithelial cell adhesion, thereby reducing the efficiency of signalling

CC through other epithelial cell receptors such as the epidermal growth

CC factor receptor. The BRAINAC and EGGHEAD proteins have cytosolic,

CC dermatological and neuroactive properties. As the BRAINAC and EGGHEAD

CC proteins are involved in adhesion between epithelial cells, and as this

CC adhesion is required for cell viability, cell growth regulation and cell

CC fate specification, it is envisioned that wild type or mutant forms of

CC mammalian BRAINAC and/or EGGHEAD can be used to alter epithelial cell

CC adhesion. This should be useful in treating many diseases which present

CC problems of cell viability, cell growth regulation and cell fate

CC specification. For example, these proteins, or active fragments or

CC analogues of these proteins and these genes can be used to treat

CC diseases such as cancer, psoriasis and other skin lesions, and nervous

CC system defects or diseases. The present sequence encodes the human

CC BRAINAC protein.

XX Sequence 1194 BP; 312 A; 242 C; 243 G; 259 T; 138 other;

SQ Query Match 60.8%; Score 1038.4; DB 21; Length 1194;

Best Local Similarity 89.5%; Pred. No. 8.3e-274;

Matches 1045; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

```
Qy 273 TTGTTGGGTATCTGTGATGATGGCAATGTCTTCATTTATTTATTTATGAGTCTCCAA 332
Db 28 TTCTGGGATCTGTGATGATGGCAATGTCTTCATTTATTTATTTATGAGTCTCCAA 87
Qy 333 AGCAGTAGCCACGAAAAAATGAAAAAGGGAGTAATATACCCAAAGAGATCTCG 392
Db 88 AACAGTAGCCACGAAAAAATGAAAAAGGGAGTAATATACCCAAAGAGATCTCG 147
Qy 393 AGATATCTACCCCTCCGAGGCAATGTAACCCAGACAGAGAGAGTGAACCGGAG 452
Db 148 AAGATATCTACCCCTCCGAGGCAATGTAACCCAGACAGAGAGAGTGAACCGGAG 207
Qy 453 TACAACCCCATCTGTGAGCATGTGACCAACACGAGCGGGAGCGGCGAGGCTCTCCAT 512
Db 208 TACAACCCCATCTGTGAGCATGTGACCAACACGAGCGGGAGCGGCGAGGCTCTCCAT 267
Qy 513 ATAAGCCATCTGAACCTACTGCGAACCTGACCTGAGGTCACGTCGCTGCTTACGGTTT 572
Db 268 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 327
Qy 573 AACAACTTCCCGACAGATTTAAAGACTTTCTGCTCTATTGAGATGCCCGCAATTATCA 632
Db 573 AACAACTTCCCGACAGATTTAAAGACTTTCTGCTCTATTGAGATGCCCGCAATTATCA 632
```


Db 343 AAGAGAAGTTCTGGAAGCCACCCAGCACATCCCGGGCATACTGGACAGGGAACAGGAGA 402
Qy 439 AGCTGAACCGGCAGTACAAACCCATCTGAGCATGCTGACCAACAGCGGGGAGGCGG 498
Db 403 AGCTGAACAGGTGTGATAATCCCATCTTGAACAGGGTGGCCAAATCAGACAGGGGAGCTAG 462
Qy 499 GCGGCTCTCCAAATATAAGCCATCTGAACCTACTCGAAGCTGACCTGAGGTCACGTCGG 558
Db 463 CCACATCTCCAAACACAGTCACTGAGCTATTGTGAACACAGATCTGACGGTCATGACAG 522
Qy 559 TGGTTAGGGTTTTAAACAATGCGGACAGATTTAAAGACTTTCTGCTGTATTGAGAT 618
Db 523 CTGTGACAGATTTTAAATAATCTGCGGACAGATTTAAAGACTTTCTGCTGTATTGAGAT 582
Qy 619 GCCCAATTATTCACTGCTATAGATCAGCGGATAGTGTGAAAAGAACCTTTCTTGT 678
Db 563 GCCGGAATTACTCGCTGCTTATAGATCAACCGAAGAAATGTGAAAAGAGCCCTTCTTAC 642
Qy 679 TCGTGGGGATTAAAGTCCCTCACTCCACATTTTGCAGAGCAAGCAATCGGGAATCCT 738
Db 643 TATTGGCGATAAGTCCCTCATTCACATTTTGCAGAGCAAGCAATTCGGGAGTCTT 702
Qy 739 GGGGCCAAGAAAGCAAGCGAGGAAACCAAGCGGTGGTGGAGTCTTCTGCTGGGCCAGA 798
Db 703 GGGGCCAGAAACCAAGCTAGGGAACACAGACAGTAGTGAGGGTCTTCTGTTGGGCAAGA 762
Qy 799 CACCCCGAGAGCAACCAACCCGACCTTTCAGATATGCTGAAATTTGAGAGTGAGAGC 858
Db 763 CACCCCGAGAGCAACCAACCCGACCTTTCGACATGCTTAAGTTTGAGAGTGACAAAGC 822
Qy 859 ACCAAGACATTCCTTATGTGAACCTACAGACACATTTCTTCAACTTGTCTCTGAAGGAAG 918
Db 823 ACCAGGACATCCTCATGTGGAACCTATAGAGACACATTTCTCAACTGTCCCTGAGGAAG 882
Qy 919 TCGTGTTCCTCAGGTGGTAAGTACTTCTCCCGACAGACATGATTTGTTTCAAGGGCG 978
Db 883 TGCTGTTTCTTAGTGGGTGAGCACTTCTCTCCAGACGCGAGAGTTTGTCTCAAGGGCG 942
Qy 979 ATGACGATGTTTTGTGAACACCCATCACATCCCTGAATTACTTCAATAGTTATCCAAAGA 1038
Db 943 ATGATGAGCTGTTGTGACACCCATCACATCCCTTAATTACTTGAATAGCTTATCCAAAGA 1002
Qy 1039 CCAAGGCCAAGATCTCTTCATAGGTGATGTGATCCCAATGCTGTGACCTCATCGGGATA 1098
Db 1003 GCAAGGCCAAGACTTGTTCATAGGTGACGTGATCCCAATGCTGTGGCTCACCGGGATA 1062
Qy 1099 AGAAGCTGAAGTACTACATCCCAAGATGTTTACTCTGGCTCAACACCTATACGAC 1158
Db 1063 AGAACTGAAGTACTACATCCCAAGATGTTTACTACACCGGCTACACCGCTGATCGCG 1122
Qy 1159 GGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGCTGTACCATATCACTGACC 1218
Db 1123 GGGGTGGTGAATCTGCTACTCCGGCCCCCTTGCTTGAGGCTGTACGTGCGACTAGCC 1182
Qy 1219 AGGTCCATCTCTACCCCATGATGACGTTTATCTGGAATGTGCTTCAGAAACTCGGCC 1278
Db 1183 GGGTCCATCTCTACCCCATGATGATGTTTATACGGGAATGTGCTTCAGAAACTCGGCC 1242
Qy 1279 TCGTTCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGGAGCAAAAACAAAATA 1338
Db 1243 TTGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATTTGAACAGAAAATAAGAAA 1302
Qy 1339 ACATCTGCTCTATGTAGATCTGATGTTAGTACATAGTAGAAAACCTCAAGAGATGATTG 1398
Db 1303 ATATTGTTCTATATAGACCTTAATGTAGTACATAGCAAAAACCTCAAGAGATGATTG 1362
Qy 1399 ATATTGTTCTCAGTTGACAGAGTCTCATTTAAATGCTAAATATAGATACAAACTCAATT 1458
Db 1363 ATATCTGGTCTCAGTTGCAAAAGTCTTAATTTAAATGCTGAATAGACATGAGCTGATT 1422
Qy 1459 TKSATWGRAGGGGTWTTTGTATWGWGCCATGTTGGGTCTCACCATTAGAGTAATT 1517
Db 1423 T--CACAGAAAGGCTAGCTGACTAGTTCCCATGGTGTGCTCTCACAATAGGTGAGTT 1479

RESULT 11
AAZ87185
ID AAZ87185 standard; cDNA; 2676 BP.
XX
AC AAZ87185.
XX
DI 08-MAY-2000 (first entry)
XX
DE Murine Brainiac cDNA.
XX
KW Brainiac; murine; mammalian; expressed sequence tag; EST;
KW Drosophila melanogaster; Egghead; Notch; epithelial cell adhesion;
KW Fringe family; cell viability; growth regulation; cell fate; cancer;
KW psoriasis; skin lesion; nervous system disorder; developmental syndrome;
KW transgenic animal; drug screening; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
CDS 210..1403
FT FT /*tag= a
FT FT /product= "Murine Brainiac protein"
PN CA2255109-A1.
XX
PD 17-JUN-1999.
XX
PF 17-DEC-1998; 98CA-2255109.
XX
PR 17-DEC-1997; 97CA-2255126.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
Egag SE:
WPL: 2000-148082/14.
P-PSDB: AAY69697.
XX
PT New nucleic acids encoding a murine and human Brainiac protein, useful
PT for detecting somatic or germline DNA lesions which are responsible for
PT developmental syndromes or diseases including cancer .
XX
PS Claim 5; Page 26-27; 40pp; English.
XX
CC This sequence represents cDNA encoding murine Brainiac protein. The cDNA
CC was isolated from a mouse mammary gland cDNA library via the use of a
CC probe generated via PCR from a variety of murine tissues. The PCR
CC primers used for probe generation (AAZ87187-287188) were based on EST
CC (expressed sequence tag) sequences with significant similarity to
CC Drosophila melanogaster Brainiac. Drosophila Brainiac and Egghead
CC proteins regulate adhesion between epithelial cells, this activity being
CC dependent on the presence of Notch. Drosophila Brainiac is a secreted
CC protein, and has sequence similarities with the Drosophila Fringe
CC for cell viability, cell growth regulation and cell fate specification.
CC Wild-type or mutant forms of mammalian Brainiac proteins may therefore be
CC used to alter epithelial cell adhesion in a mammal. Mammalian Brainiac
CC proteins, active fragments analogues, and nucleic acids may be used to
CC treat diseases such as cancer, psoriasis and other skin lesions, and
CC nervous system disorders. Mammalian Brainiac nucleic acids may also be
CC used to detect somatic or germline DNA lesions which are responsible for
CC developmental syndromes or diseases including cancer. The mammalian
CC Brainiac proteins and fragments or its analogues are useful as antigens
CC in immunoassays including enzyme-linked immunosorbent assays (ELISA),
CC radioimmunoassays (RIA) and other non-enzyme linked antibody binding
CC assays. Non-human transgenic animals comprising nucleotide sequences
CC encoding human Brainiac protein(AAY69698) can be used as animal models
CC for the study of mammalian Brainiac gene function, for the screening of
CC candidate compounds and for the evaluation of potential therapeutic
CC interventions.
XX

SQ Sequence 2676 BP; 732 A; 562 C; 637 G; 744 T; 1 other;

Query Match	58.7%;	Score 1002.2;	DB 21;	Length 2676;
Best Local Similarity	83.0%;	Pred. No. 1.le-263;		
Matches 1195; Conservative	7;	Mismatches 225;	Indels 12;	

Qy	79	CGCCGACACCGCCGGCCGCCCTTCGCGGGGGCGGCGCATGAGAGCTGAGCTGGCGCGGT	138
Db	53	CGCAGCCCGCCGGCCGCCCGTCC-GGCCATTGGCGCATGGAGCGAGAG-CGCGCGGT	110
Qy	139	CGCCGGCTGAGCCGCGGAGCGGGCGGACGTGGATGTGGCCGCGATCTCCCGCTT	198
Db	111	CGCGGGCTGAGCGCAK-GACCGGCTGGACGTGGATCGCGCGCGGTCTTCGG-----	164
Qy	199	GCCTCCGCCCGCCGAGCTGAGGTGCTCCGGGACAGATATGAGAAATGAGTGTGGAC	258
Db	165	--CCCGCCCGCCGAGCTGGAGTGTCCCTAGACAGGTATGAGAGATGAGCTGGGGC	222
Qy	259	GTCAAGAATAAAGTTGTTGGTATCTCTGATGTGGCAATGCTTCATTTATTTATTA	318
Db	223	GTCAAGAGTGAAGTTGCTGGCATCTGATGGCAATGCTTCATTTATTTGATG	282
Qy	319	TGGAGTCTCCAAAACGAGTAGCCACGAAANAATGGAAAAGGGGAAGTAATAATACCA	378
Db	283	TGGAGTCTCCAAAACAGTAGCCACGAAANAATGGAAAAGGGAGGTAATAATCCGA	342
Qy	379	AACAGAAGTCTGGAAGATATCTACCCCTCCGAGGCATACTCGAACCCAGAGCAAGAG	438
Db	343	AAGAGAAGTCTTGAAGCCACCAGCACTCCCGGGCATACTGGACAGGGAACACGAGA	402
Qy	439	AGCTGAACCGGAGTACAAACCCATCTCTGAGCATGCTGACCAACCAGACGGGGAGRCGG	498
Db	403	AGCTGAACAGTGTGTAATCCCATCTGTGAACAGGCTGGCCAATCAGACAGGGAGCTAG	462
Qy	499	GCAGGCTCTCCAATTAAGCATCTGAACTACTCGAACCTGACCTCAGGCTCAGCTCGG	558
Db	463	CCACATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACTCGAGGCTCATGACAG	522
Qy	559	TGCTTACGGGTTTTAACACTTGGCGGACAGATTTAAAGACTTCTGCTGTATTTGAGAT	618
Db	523	CTGTGACAGATTTAATAATCGCGGACAGATTTAAAGACTTCTGCTGTATTTGAGAT	582
Qy	619	CGCCGAATTTACTCTGCTTATAGATCAGCGGATAGTGTGCAAGAAACCTTCTCTGT	678
Db	583	CGCGAATTTACTCTGCTTATAGNTCAACGGAAGAATGTGCAAGAACCCCTCTTAC	642
Qy	679	TGCTGGCGATTAAGTCCCTCACTCCATTTTGGCCAGAGGCAAGCAATCCGGGAATCT	738
Db	643	TATTTGCGATAAAGTCCCTCAATCCATTTTGGCCAGAGGCAAGCAATCCGGGAGCT	702
Qy	739	GGGGCCCAAGAAACACGACAGGAACCAACGGTGGTGGAGTCTTCTCTGCTGGGCGACA	798
Db	703	GGGGCCGCAAAACCACTAGGAACACAGACAGTAGTGGGTCTTCTCTTGGGCAAGA	762
Qy	799	CACCCCCAGAGCAACCAACCCGACCTTTCAGATATGCTGAAATTTGAGAGTGAGAGCC	858
Db	763	CACCCCCAGAGCAACCAACCCGACCTTTCGAGATGCTTAAGTTTGAGAGTGACAC	822
Qy	859	ACCAAGACATCTTATGTTGAACCCCATCATCTCTGAATTTACTTGAATAGTTTATCCA	918
Db	823	ACCAGGACATCTCATGTGGAACTATAGACACATTTCTCAACCTGTCCCAGAGGAG	882
Qy	919	TGCTGTTTCTCAGTGGGTAGTACTTCTCCACAGACACAGTGTGTTTCAAGGGCG	978
Db	883	TGCTGTTTCTCAGTGGGTAGCACTTCTCTCCAGACGAGAGTGTGTTTCAAGGGCG	942
Qy	979	ATGACGATGTTTGTGAAACCCCATCATCTCTGAATTTACTTGAATAGTTTATCCA	1038
Db	943	AIGATGAGCTGTTTGAACACCCCATCATCTCTTAATTTACTTGAATAGTTTATCCA	1002
Qy	1039	CCAAAGCCAAAGATCTTTCATAGGTGATGATCCACAAATGCTGGACCCFATCGGGATA	1098
Db	1003	GCAAAGCCAAAGACTTGTTCATAGTGAAGTGAATCCCAATGCTGGCCCTACCGGGATA	1062

RESULT 12

ABV75081
ID ABV75081 standard; DNA: 1260 BP.

XX AC ABV75081;

XX DT 19-FEB-2003 (first entry)

XX
DE Murine beta3Gnt gene polynucleotide sequence.

XX Beta3GnT: transgenic; beta-1,3-N-acetylglucosaminyltransferase;
KW disease system: gene; mouse; mouse; ds.

XX	Mus sp.
OS	

XX	
FH	Key
FT	CDS
FT	Location/Qualifiers 30..1223 /*tag= a /product= "beta3Cn"

AA WO200279413-A2
PN

XX
PD
10-OCT-2002

XX
PF 29-MAR-2002: 2002WO-US09645.

XX
PR 29-MAR-2001: 2001US-280706P.

PR 28-MAR-2002; 2002US-0280706.
YX

PA (DELT-) DELTAGEN INC.

PI Leviten MW, Phillips R;

DR WPI; 2003-067437/06.

UK F-F3DB; ABB62040.
XX

PT New transgenic mouse
PT vivo model to study v

PT beta3GnT may be impli

XX

XX
XX

XX PS Claim 16; Fig 1A-B; 123pp; English.

XX CC The present sequence encodes a leptin induced LIG46 polypeptide. The

XX CC specification describes a method for determining whether a compound can

XX CC be used to modulate body weight by measuring the activity of leptin

XX CC inducible genes, such as LIG46. The method can be used to specifically

XX CC detect LIG46 and LIG56 nucleic acid molecules relative to other nucleic

XX CC acid molecules encoding galactosyltransferases or GTP-binding proteins.

XX CC They can also be used in diagnostic assays to identify the presence or

XX CC absence of a genetic lesion or mutation characterized by aberrant

XX CC modification or mis-regulation of the genes or aberrant post-

XX CC translational modification of the proteins. LIG46 and LIG56 proteins

XX CC and nucleic acid molecules can be used to treat obesity or cachexia.

XX CC LIG46 and LIG56 antagonists are used to treat obesity and LIG46 and

XX CC LIG56 agonists are used to treat low body weight. The leptin inducible

XX CC genes i.e. LIG46, LIG56, Tgtp, LRG-47, RC10-II and Stral3 can be used

XX CC to determine whether a compound modulates body weight and can then be

XX CC used to treat obesity or cachexia or low body weight.

XX SQ Sequence 1196 BP; 350 A; 269 C; 290 G; 287 T; 0 other;

Query Match 53.1%; Score 906.4; DB 21; Length 1196;

Best Local Similarity 84.9%; Pred. No. 1.2e-237;

Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 244 AATGAGCTGTTGGAGTCGGAAGATAAAGTTGTTGGGTATCTTCCTGATGATGGCAATGTCT 303

DB 1 AGATGAGTGTGGGGCGTCGAAGAGTCAAGTTGCTGGGCATCCCTGATGATGGCAATGTCT 60

QY 304 TCATTATTTTATTATTAAGCTCTCCAAAGACAGTCCAGCAAAATAATGGAAGGGG 363

DB 61 TCATTATTTGATGTCGAGTCTCCAAAGACAGTCCAGCAAAATAATGGAAGGGG 120

QY 364 AAGTAATATACCAAGACAGAGTTCTGGAGATATCTACCCCTCCGAGGCACTACTGGA 423

DB 121 GAGTAATATCCGAAGAGAGAGTTCTGGAGGCCACCCAGCAGTCCCGGGGCACTACTGGA 180

QY 424 ACCGAGACCAAGAGAGTGAACGGGAGTACACCCCATCTGAGCATGCTGACCAACC 483

DB 181 ACAGGACACAGGAGAGTGTACAGAGTGTGTACATCCATCTTGAACAGGGTGGCCAAATC 240

QY 484 AGACGGGGAGGGGGAGGCTCTCCATATATAGCCATCTGAACTACTGCGAAGCTGACC 543

DB 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAAAGTCACTGAGCTATTTGTGAACAGACT 300

QY 544 TGAGGGTCACGTCGGTGTGTACGGGTTTAAACACTTGGCGGACAGATTTAAAGACTTTC 603

DB 301 CGACGGTCATGACAGCTGTGACAGATTTAATATCTGCGGACAGATTTAAAGACTTTC 360

QY 604 TGCTGTATTGAGATGCGGCAATATTACTGCTTATAGATFACCGCGGATAGTGTGCAA 663

DB 361 TCTTGTATTGAGATGCGGCAATATTACTGCTTATAGATFACCAAGAAATGTGCAA 420

QY 664 AGAACCCTTCTTGTGTGGGATTAAGTCCCTCACTCCACATTTTGCAGAGAGGCAAG 723

DB 421 AGAAGCCCTTCTTGTGTGGGATTAAGTCCCTCACTCCACATTTTGCAGAGAGGCAAG 480

QY 724 CAATCGGGGAATCTTGGGGCAAGAACCAACGAGGGAACCAACGGTGTGCGAGTCT 783

DB 481 CAATTCGGGAGTCTTGGGGCGGAGAACCAACGAGTGGGAACACAGACTAGTGGGTCT 540

QY 784 TCTCTGTGGGCGACACACCCCGAGAGGACACACCCCGGACCTTTCAGATATGCTGAAT 843

DB 541 TCTCTGTGGGCAAGACACCCCGAGAGGACACACCCCGGACCTTTCGACATGCTTAAGT 600

QY 844 TTGAGAGTCAGACACCAAGACATCTTATGTGGAATAGACACACTTCTTCAACT 903

DB 601 TTGAGAGTCAGACACCAAGACATCTTATGTGGAATAGACACACTTCTTCAACC 660

QY 904 TGTCTCTGAAGGAAGTGTGTTCTTCAGGTGGTAAAGTACTTCTGCGGACACACTGAGT 963

DB 661 TGTCTCTGAAGGAAGTGTGTTCTTAGTGGGTGAGCACTTCTGTCACACCGACAGT 720

QY 964 TTGTTTTCAGGGCGATGACAGATGTTTTTGTGAACACCCATCATCATCTCTGAATTACTTGA 1023

DB 721 TTGTTTTCAGGGCGATGACAGTGTGTTGTGAACACCCATCATCATCTCTGAATTACTTGA 780

QY 1024 ATAGTTTATCCAAAGACCAAGCCAAAGATCTCTTCATAGGTGATGTGATCCCAATGCTG 1083

DB 781 ATAGCTTATCCAAAGACCAAGCCAAAGACTTGTTCATAGGTGAGCTGATCCCAATGCTG 840

QY 1084 GACCTCATCGGGATGAAGACTGAAGTACTACTATCCAGAAAGTTGTTTACTCTGGCTCT 1143

DB 841 GGCCTCACCGGATGAAGAACTGAAGTACTACTATCCAGAAAGTCTTCTACACCGGCTCT 900

QY 1144 ACCACCCCTATCCAGGGGAGGGGGTTCCTCTACTCCGGCCACTGCGCTGAGGCTGT 1203

DB 901 ACCCACCCTATCCGGGGGTGGTGAATTCCTGTTACTCCGGCCCTTGCCTTGAGGCTGT 960

QY 1204 ACCATATCATCTGACGAGCTCCATCTCTACCCCATTTGATGAGCTTTATATGGAATGSCC 1263

DB 961 ACAGTGGCACTAGCGGGTCCATCTCTACCCCTATTGATGATGTTTATACGGGAATGTC 1020

QY 1264 TTCAGAAACTCGGCTCGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCCAGG 1323

DB 1021 TTCAGAAACTCGGCTCGTTCCAGAGAAAGCACAAGGCTTCAGGACATTTGATATGAAG 1080

QY 1324 AGAAAAACAAAAATAACATCTGCTCTATGTAGATCTGATGTTAGTACATAGTAAAAAC 1383

DB 1081 AGAAAAATAAGAAAAATATTGTTCTCTATATAGACCTAATGTTAGTACATAGCAGAAAC 1140

QY 1384 CPCAAGAGATGATGATATTGTTGGTCTCAGTTCAGAGTGCAGAGTGCATTTTAAATGCTAA 1439

DB 1141 CTCAAGAGATGATGATATATCTGCTCAGTTCGAAAGTCTTAATTTAAATGCTGA 1196

RESULT 14

AA167865

ID AA167865 standard; cDNA; 1196 BP.

XX AC AA167865;

XX AC AA167865;

DI 13-MAR-2002 (first entry)

DE Murine LIG46 polypeptide encoding cDNA.

XX KW Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Stral3;

XX KW anorectic; anabolic; antisense therapy; mouse; ss.

XX OS Mus sp.

XX Key Location/Qualifiers

FT CDS 3..1194

FT /tag= a

FT /product= "LIG46"

FT /note= "the coding sequence is specifically claimed"

FT sig_peptide 3..98

FT /tag= b

FT mat_peptide 99..1193

FT /tag= c

XX US2001024808-A1.

PN 27-SEP-2001.

XX 12-MAR-2001; 2001US-0804357.

XX 29-OCT-1998; 98US-106378P.

XX 19-NOV-1998; 98US-D195896.

XX 10-SEP-1998; 98US-0150857.

XX (MILL-) MILLENNIUM PHARM INC.

XX White D, Zhou J, Tartaglia LA;

DR WPI: 2001-624963/72.
DR P-PSDB: AAG66115, AAG66116.
XX An isolated polypeptide useful for treating a weight disorder and for
PT screening compounds that may inhibit protein activity comprises a
PT leptin-induced protein
XX
PS Claim 16; Fig 1; 46pp; English.
XX
CC The invention relates to genes whose expression are induced by leptin.
CC LIG46 and LIG56 are novel leptin induced genes (LIG), while four other
CC genes such as LRP-47, RC10-II and Stral3 have been previously
CC identified. The leptin induced proteins can be expressed by standard
CC recombinant methodology. The proteins and encoding polynucleotides may
CC be used in screening assays to identify compounds that may bind to it.
CC Administering a molecule, e.g., an antisense molecule, which reduces
CC expression of activity of protein selected from LIG46, LIG56, LRP-47,
CC LRP-47, RC10-II, and Stral3 may be used to treat a weight disorder. The
CC method may further comprise administering leptin. The present sequence
CC represents a cDNA encoding a murine LIG46 protein.
XX
XX Sequence 1196 BP; 350 A; 269 C; 290 G; 287 T; 0 other;

Query Match 53.1%; Score 906.4; DB 23; Length 1196;
Best Local Similarity 84.9%; Pred. No. 1.2e-237;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 244 AATGAGTGTGGACGTCGAGAGATAAAGTTGTGGTATCTCTGATGGGCAATGCT 303
DB 1 AGATGAGTGTGGGCGGTGGAAGATCAAGTTGCTGGGCATCTCTGATGGCAATGCT 50
QY 304 TCATTTATTTATATGGAAGTCTCCAAAGCAGTAGCCAAAGAAAATGGAAGGGG 363
DB 61 TCATTTATTTGAGTGTGGAAGTCTCCAAAGCAGTAGCCAAAGAAAATGGAAGGGG 420
QY 364 AAGTAATAATACCAAGAGAGTTCTGGAAGATATCTACCTCCCGAGGCACTGGA 423
DB 121 GACTAATAATCCCGAAAGAGAGTTCTGGAAGCAGCCAGCAGCTCCCGGCGCATCTGGA 180
QY 424 ACCGAGAGCAGAGCAAGCTGAACCGGCGAGTAGACACCCATCTGAGCAGTCCGACCAAC 483
DB 181 ACAGGAGCAGGAGCAAGCTGAACCGGCGAGTAGACACCCATCTGAGCAGTCCGACCAAC 440
QY 484 AGACGGGGGAGCGGCGAGCTCTCCAAATATAGCCATCTGAACCTGCGAACCCTGACC 543
DB 241 AGACGGGGAGCTAGCCACATCTCCAAACACAGTCACCTGAGCTATTTGTGACACAGACT 309
QY 544 TGAGGGTCAGCTCGGTGGTTACGGGTTTAAACACITGCGGACAGATTTAAGACTTTC 603
DB 301 CGACGGTCATGACAGCTGTGACAGATTTTAAATATCTGCGGACAGATTTAAGACTTTC 360
QY 604 TGCTGATTTGAGATCGCGCAATTAATCACTGCTTATAGATCAGCGGATAAGTGTCAA 663
DB 361 TCTGTATTTGAGATCGCGCAATTAATCACTGCTTATAGATCAGCGGATAAGTGTCAA 420
QY 664 AGAAACCTTTCTTTGCTGCGGATTAAGTCCCTCACTCCACATTTTCCAGAGGCAAG 723
DB 421 AGAAGCCCTCTTACTATTGGCGATAAGTCCCTCACTTCCACATTTTCCAGAGGCAAG 480
QY 724 CAATCCGGGATCTCGGGCCAGAAAGCAAGCAGGAGCAACCAAGCGTGGTCCGAGTCT 783
DB 481 CAATCCGGGATCTCGGGCCAGAAAGCAAGCAGGAGCAACCAAGCGTGGTCCGAGTCT 540
QY 784 TCCTGCTGGGCGACACACCCCGAGGAGCAACCAACCCCGACCTTTTCAGATATGCTGAAT 843
DB 541 TCCTGTTGGGCAAGACACACCCCGAGGAGCAACCAACCCCGACCTTTTCAGATATGCTGAAT 600
QY 844 TTGAGAGTGAGAGCAGCAGCAAGATTTCTTATGTGGAAGTACAGAGACACTTTCTCAACT 903
DB 601 TTGAGAGTGAGAGCAGCAGCAGCAAGATTTCTTATGTGGAAGTACAGAGACACTTTCTCAACT 660
QY 904 TGTCTCTGAAGAGTCTGTTCTCAGGTGGGTAACTCTCTCCAGAGCACTGACT 963
DB 904 TGTCTCTGAAGAGTCTGTTCTCAGGTGGGTAACTCTCTCCAGAGCACTGACT 963

DB 661 TGCCCTGAAGGAAGTCTCTGTTCTTAGTGGGTGAGCACITCTCTCCAGAGCGAGAGT 720
QY 954 TTGTTTTCAGGGCGATGACGATGTTTGTGAACACCCCATCACATCTCTGAACTACTTGA 1023
DB 721 TTGTTTTCAGGGCGATGACGATGTTTGTGAACACCCCATCACATCTCTGAACTACTTGA 780
QY 1024 ATAGTTTATCAGAGCAACCAAGCATCTTTCTATAGTGTGATGATCCCAATGCTG 1083
DB 781 ATAGCTTATCAGAGCAACCAAGCATCTTTCTATAGTGTGATGATCCCAATGCTG 840
QY 1084 GACCTATCGGGATTAAGAAGCTGAAGTACTACATCCCAAGAGTGTGTTACTCTGGCCTCT 1143
DB 841 GGCCTACCGGGATTAAGAAGCTGAAGTACTACATCCCAAGAGTCTTCTACACCGGCTCT 900
QY 1144 ACCCACCCTTATGACAGGGGAGGGGGTTCCTTACTCCGCGCACCTGGGCGCTGAGGCTGT 1203
DB 901 ACCCACCCTTATGACAGGGGAGGGGGTTCCTTACTCCGCGCACCTGGGCGCTGAGGCTGT 960
QY 1204 ACCATATCACTGACAGGTCCTCTCTACCCCATTTGATGACGTTTATCTGGAATGCTGC 1263
DB 961 ACAGTGGGACTAGCGGGTCTCTCTACCCATTTGATGATGTTTATACGGGANTGCTGC 1020
QY 1264 TTCAGAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGACATTTGATATCGAGG 1323
DB 1021 TTCAGAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGACATTTGATATCGAGG 1080
QY 1324 AGAAACACAAATTAACATCTGCTCTCTATGATGATGATGATGATGATGATGATGATGAT 1383
DB 1081 AGAAACACAAATTAACATCTGCTCTCTATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1384 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1439
DB 1141 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1196

RESULT 15
AAD46672
ID AAD46672 standard; cDNA; 1196 BP.
XX AC AAD46672;
XX AC AAD46672;
DT 27-JAN-2003 (first entry)
XX Murine LIG46 cDNA.
XX LIG46; leptin; low body weight; chromosomal mapping; tissue typing;
KW forensic biology; transgenic; gene therapy; antianorectic; murine;
KW gene; ss.
OS Mus sp.
XX Key
FH 3..1196 Location/Qualifiers
FT CDS
FT /*tag= a /product= "Murine LIG46 protein"
FT sig_peptide 3..98
FT /*tag= b
FT mat_peptide 99..1193
FT /*tag= c
FT /product= "Murine mature LIG46 protein"
XX WO200274905-A2.
PN 26-SEP-2002.
XX 20-NOV-2001; 2001WO-US43345.
XX 21-NOV-2000; 2000US-0717778.
XX (MILL-) MILLENNIUM PHARM INC.
XX White DW, Zhou J, Tartaglia LA, Stricker-krongrad A, Clausen H;
XX

DR WPI; 2002-759886/82.
DR P-PSDB; AAE29092.
XX
XX New isolated nucleic acid molecules, designated as leptin-induced genes
PT 46 (Lig46), useful for treating a subject having a disorder
PT characterized by undesirable level of Lig46 expression or activity,
PT such as low body weight -
XX
XX Claim 15; Fig 1; 90pp; English.
XX
XX The invention relates to Lig46, a gene whose expression is induced by
CC leptin. Lig46 DNA and protein are useful in treating a subject having
CC a disorder characterized by undesirable level of Lig46 expression or
CC activity, such as low body weight. They are also useful in a screening
CC assay, chromosomal mapping, tissue typing and forensic biology. The
CC probes based on the Lig46 nucleotide sequence are useful for detecting
CC transcripts or genomic sequences encoding the same or related proteins.
CC The Lig46 polypeptides are useful as immunogens for raising anti-Lig46
CC antibodies. The host cells are useful for producing non-human transgenic
CC animals. Lig46 DNA is used in gene therapy. The present sequence is
CC murine Lig46 cDNA.
XX
SQ Sequence 1196 BP; 350 A; 269 C; 290 G; 287 T; 0 other;
Query Match 53.1%; Score 906.4; DB 24; Length 1196;
Best Local Similarity 84.9%; Pred. No. 1.2e-237;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
QY 244 AATGAGTGTGGACGTCGAGGAATAAAGTTGTGGGTATCTCGATGATGGCAATGTCT 303
DB 1 AGATGAGTGTGGGCGCTCGAAGAGTCAAGTGTCTGGGCATCTCGATGATGGCAATGTCT 60
QY 304 TCATTTATTTATTTATGGAAGTCTCCAAAGAGTAGCCCAAGAAAAAATGGAAGGGG 363
DB 61 TCATTTATTTATTTGGAAGTCTCCAAAGAGTAGCCCAAGAAAAAATGGAAGGGAG 120
QY 364 AAGTAATAATACCAAGAGAGTCTCGAAGATATCTACCCCTCCGAGGCGACTACTGA 423
DB 121 GAGTAATAATCCCGAAAGAGAGTCTGGAAGCAGCCAGCACTCCCGGGCATACTGGA 180
QY 424 ACCGAGAGCAAGAGCACTGAACCGGCGAGTACAAACCCCTCTCGAGCATGCTGACCAACC 483
DB 181 ACAGGGANACAGGAGAGCTGAACAGGTGTTGACATCCCATCTTGAACAGGTGGCCAATC 240
QY 484 AGACGGGGAGGGGGGCGAGCTCTCCAAATATAGCCATCTGAACACTACTCGCAACCTGACC 543
DB 241 AGACAGGGGAGCTAGCCACATCTCCAAACCAAGTCACTGAGCTATTGTGAACAGAGCT 300
QY 544 TGAGGGTCACGTCGCTGTTACGGGTTTAAACACTTGGCGGACAGATTTAAGACTTTC 603
DB 301 CGAGCGTCATGACAGCTGTGACAGATTTTAAATCTGCGGACAGATTTAAGACTTTC 360
QY 604 TGCTGTATTTGAGATGCGGCAATATTACTGCTTATAGATCAGCCGGATAGGTGTGCAA 663
DB 361 TCTGTATTTGAGATGCGGCAATATTACTGCTGCTATAGATCAACCGAAGAAATGTCAA 420
QY 664 AGAANCTTCTGCTGGCGATAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 723
DB 421 AGAAGCCCTCTTACTATTGGCGATAAAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
QY 724 CAATCCGGGAATCTGGGCGCAAGAGCAACCGGAGCAACCGGAGTGGTGGCGAGTCT 783
DB 481 CAATTCGGGAGCTTGGGCGGAGAAACCAACCGTAGGGAACAGAGAGTGTGAGGTCT 540
QY 784 TCTGTGTTGGGCGGAGACACCCCGGAGGACCAACCCCGGAGTGTGAGTGTGAGTGTGAGT 843
DB 541 TCTGTGTTGGGCGGAGACACCCCGGAGGACCAACCCCGGAGTGTGAGTGTGAGTGTGAGT 600
QY 844 TTGAGGTGAGAGCAACCAAGATCTTATGTGGAACTACAGAGACTTTCTTCAACT 903
DB 601 TTGAGGTGAGAGCAACCAAGATCTTATGTGGAACTATAGAGACATTTCTTCAACC 660
QY 904 TGTCTCTGAAGGAGTGTCTGTTTCTCAGGTGGGTAAAGTACTTCTCTGCCAGAGACTGAGT 963

DB 661 TGTCCCTGAAGAGTGTCTGTTCTTAGGTGGTGAGCACTTCTCTCCAGACGAGAGT 720
QY 964 TTGTTTCAAGGGCGATGACGATGTTTGTGAACACCCATCACATCTCTGAATTAATTGTA 1023
DB 721 TTGTTTCAAGGGCGATGACGATGTTTGTGAACACCCATCACATCTCTGAATTAATTGTA 780
QY 1024 ATAGTTTATCCAAAGACCAAGCAAGATCTCTTCATAGGTGATGTGATCCACAAATGCTG 1083
DB 781 ATAGCTTATCCAAAGACCAAGCAAGACTTCTTCATAGGTGATGTGATCCACAAATGCTG 840
QY 1084 GACTCATCGGGATACAAAGCTGAAGTACTATCCAGAGAGTGTGTTTACTCTGGCTCT 1143
DB 841 GGCTCATCCGGGATACAAAGCTGAAGTACTATCCAGAGAGTGTGTTTACTCTGGCTCT 900
QY 1144 ACCACCTATGACGGGGAGGGGTTCTCTACTCTCCGGCCACCTGGCCCTGAGGCTGT 1203
DB 901 ACCACCTATGACGGGGAGGGGTTCTCTACTCTCCGGCCACCTGGCCCTGAGGCTGT 960
QY 1204 ACCATATCACTGACAGGTCCATCTCTACCCATTGATGACGCTTTTACTGGAATGTGCC 1263
DB 961 ACAGTGGCACTAGCCGGGTCATCTCTACCCATTGATGATGATGTTTATACGGGAATGTGCC 1020
QY 1264 TTCAGAAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGG 1323
DB 1021 TTCAGAAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATGAG 1080
QY 1324 AGAAACAAATAATACATCTCTCTATGTAGATCTGATGTTAGTACATAGTAAAC 1383
DB 1081 AGAAATAAGAAAAATATTTGTTCTATATAGACCTAATGTAGTACATAGCAGAAAC 1140
QY 1384 CTCAGAGATGATGATTTGTTCTCAGTTGCGAGGTGCTCATTTAAATGCTAA 1439
DB 1141 CTCAGAGATGATGATTTGTTCTCAGTTGCGAGGTGCTCATTTAAATGCTAA 1196

Search completed: October 20, 2003, 14:46:01
Job time : 488.67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 14:27:23 : Search time 3857.36 Seconds
(without alignments)
10755.472 Million cell updates/sec

Title: US-09-804-357B-13

Perfect score: 1707

Sequence: 1 acgcgtccgcgcagcggcag.....aatctgttgatggccctt 1707

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	899.2	52.7	1201	9 AL553132	AL553132
2	717.2	42.0	832	9 AUI137528	AUI137528
3	708.2	41.5	876	13 BX370083	BX370083
4	659.6	38.6	816	12 B1546809	B1546809

5	556.8	38.5	743	14	CB554366
6	651.2	38.1	692	10	BG106243
7	648.6	38.0	1053	12	BM557603
8	624.4	36.6	725	12	B1549804
9	573.2	33.6	1432	12	BM926933
10	565	33.1	659	10	BG741357
11	530	31.0	632	10	BG721682
12	508.6	29.8	680	14	CB457941
13	488.2	28.6	701	9	AV724042
14	463.6	27.2	608	10	BF001654
15	458.8	26.9	593	10	AW958548
16	433.2	25.4	558	12	BM537435
17	431.4	25.3	857	14	W26453
18	422.4	25.2	547	14	CD560301
19	422.4	24.7	945	13	B0369620
20	414.4	24.3	587	10	BE623172
21	410	24.0	801	10	BE622082
22	408.6	23.9	609	10	BE100018
23	407.8	23.9	436	9	AV732823
24	406.6	23.8	589	9	AW495241
25	404.4	23.7	1084	10	BE894128
26	399.6	23.4	592	10	BF467111
27	398.8	23.4	424	9	AV682596
28	397	23.3	865	13	B0427178
29	395.4	23.2	686	14	BY754593
30	393.6	23.1	505	9	AZ04363
31	390.8	22.9	664	14	BY724171
32	385.4	22.6	655	14	BY753386
33	379.4	22.2	517	9	AA119132
34	370.2	21.7	488	10	BE632125
35	366.8	21.5	476	9	AA986646
36	364.8	21.4	457	9	AA733556
37	358.2	21.0	543	9	AW530344
38	345	20.2	610	10	BB619419
39	341	20.0	964	13	BQ893660
40	335.6	19.7	501	10	BE685687
41	334.8	19.6	490	9	A1408668
42	334.2	19.6	489	10	BC062994
43	328.2	18.2	480	14	H78875
44	326.8	19.1	410	10	BF469505
45	326.8	19.1	484	9	A1407754

ALIGNMENTS

RESULT 1
AL553132
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AL553132 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA clone CS0DI074YK14 5-PRIME, mRNA sequence.
AL553132 1201 bp mRNA linear EST 31-MAY-2003
AL553132 Homo sapiens (human)
Homo sapiens
Fukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12892683.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 1713.f For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DI074F07QPl&cluster=1713.f. Contact : Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 160C
Faraday Avenue Genoscope sequence ID : CS0D1074BF07QPL

FEATURES

Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1074YK14"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT
ORIGIN

Query Match 52.7%; Score 899.2; DB 9; Length 1201;
Best Local Similarity 98.5%; Pred. No. 3.7e-196;
Matches 954; Conservative 5; Mismatches 4; Indels 6; Gaps 5;

Qy	9	CGCAGCGGCGGAGCGGACAGCGGACAAAGTCCGGAGGCTAGCAGACCCAAAGCGCGAG	68
Db	80	CGCACGCGGAGCGGAGCGGACAGCGGACAAAGTCCGGA-GTAGCAGACCCAAAGCGCGAG	138
Qy	69	CAGTCCCTGCGCGGACCGCGGCGCGCGCTCCGGGCGCGCGCATGGAGCGTGAG	128
Db	139	CAGTCCCTGCGCGGACCGCGGCGCGCGCTCCGGGCGCGCGCATGGAGCGTGAG	198
Qy	129	CTCGGGCGGTCGCGGCGTGAAGCGGCGGAGCGCGGAGCGTGGATGTGCGCCGCGATC	188
Db	199	CTCGGGCGGTCGCGGCGTGAAGCGGCGGAGCGCGGAGCGTGGATGTGCGCCGCGATC	256
Qy	189	TCGCGGCTTGCCTGCGCGGCGGAGCTGGAGCTGCTCCGACAGATATGAGAAATG	248
Db	257	TCGCGGCTTGCCTGCGCGGCGGAGCTGGAGCTGCTCCGACAGATATGAGAAATG	316
Qy	249	AGTGTGGAGCTGCAAGATTAAGTGTGGGTATCTGATGATGCAAAATGCTTCATT	308
Db	317	AGTGTGGAGCTGCAAGATTAAGTGTGGGTATCTGATGATGCAAAATGCTTCATT	376
Qy	309	TATTTATTATGGAAGTCTCCAAAGCAGTAGGCAAGAAATAATGGAAAGGGGAAGTA	368
Db	377	TATTTATTATGGAAGTCTCCAAAGCAGTAGGCAAGAAATAATGGAAAGGGGAAGTA	436
Qy	369	ATAATACCAAGAGAGTCTCGAAGATATCTACCCCTCCGAGGCATCTGGAACCGA	428
Db	437	ATAATACCAAGAGAGTCTCGAAGATATCTACCCCTCCGAGGCATCTGGAACCGA	496
Qy	429	GAGCAAGAGAGTGAACCGGCGAGTACACCCCATCTGAGCATGCTGACCAACGAGC	488
Db	497	GAGCAAGAGAGTGAACCGGCGAGTACACCCCATCTGAGCATGCTGACCAACGAGC	556
Qy	489	GGGAGGCGGCGAGGCTCCATATATAGCCATCTGAAGTCTGCAAGCTGACCTGAG	548
Db	557	GGGAGGCGGCGAGGCTCCATATATAGCCATCTGAAGTCTGCAAGCTGACCTGAG	616
Qy	549	GTCAGTCCGTGTTACGAGTTTAAACAGTTCGCGGAGAGATTTAAAGACTTCTGCTG	608
Db	617	GTCAGTCCGTGTTACGAGTTTAAACAGTTCGCGGAGAGATTTAAAGACTTCTGCTG	676
Qy	609	TATTTGAGATCCGCAATTTATCTGCTTATAGATCAGCCGATTAAGTGTGCAAGAA	568
Db	677	TATTTGAGATCCGCAATTTATCTGCTTATAGATCAGCCGATTAAGTGTGCAAGAA	736
Qy	669	CTTTCTTGTGCTGGGATTAAGTCCCTCACTCCACATTTTCCAGAGGCAAGCAATC	728
Db	737	CTTTCTTGTGCTGGGATTAAGTCCCTCACTCCACATTTTCCAGAGGCAAGCAATC	796
Qy	729	CGGAATCTCTGGGCGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG	788
Db	797	CGGAATCTCTGGGCGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG	855

Qy	789	CTGGCCAGACACCCCGAGGAGCAACCCCGGACCTTTCAGATATGCTGAAATTTGAG	848
Db	856	CTGGCCAGACACCCCGAGGAGCAACCCCGGACCTTTCAGATATGCTGAAATTTGAG	915
Qy	849	AGTGAGAAGCACCAGACATTTTATGTGGAACCTACAGAGACACTTTCTTCAACTTGTCT	908
Db	916	AGTGAGAAGCACCAGACATTTTATGTGGAAGCTACAGAGACACTTTCTTCAACTTGTCT	975
Qy	909	CTGAAGGAAG-TGCTGTTTCTCAGTGGGTAAAGTACTTCTGCCAGACACTGAGTTGT	967
Db	976	CTGAAGGAAGTGTGTTTCTCAGTGGGTAAAGTACTT-CTGCCAGACACTGAGTTGT	1034
Qy	968	TTTCAAGGG 976	
Db	1035	TTTCAAGGS 1043	

RESULT 2
AUI37528
LOCUS AUI37528 832 bp mRNA linear EST 02-AUG-2002
DEFINITION AUI37528 PLACE1 Homo sapiens cDNA clone PLACE1006678 5', mRNA
sequence.
ACCESSION AUI37528
VERSION AUI37528.1 GI:10998067
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 832)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yanamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1..832
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1006678"
/tissue_type="placenta"
/clone_lib="PLACE1"
/note="vector: pME18SFL3"

FEATURES

source
Location/Qualifiers
1..832
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1006678"
/tissue_type="placenta"
/clone_lib="PLACE1"
/note="vector: pME18SFL3"

BASE COUNT 248 a 193 c 194 g 194 t 3 others

ORIGIN

Query Match 42.0%; Score 717.2; DB 9; Length 832;
Best Local Similarity 96.2%; Pred. No. 2.8e-154;
Matches 787; Conservative 0; Mismatches 25; Indels 6; Gaps 5;

Qy	320	GGAAGTCTCCAAAAGCAGTAGCCCAAGAAAAAATGAAAAAGGGGAAGTAATAATACCCAA	379
Db	9	GGAAGTCTCCAAAAGCAGTAGCCCAAGAAAAAATGAAAAAGGGGAAGTAATAATACCCAA	68
Qy	380	AGAGAAGTCTGGAAGATATCTACCCCTCCGAGGCATCTGGAACCGAGAGCAAGAA	439
Db	69	AGAGAAGTCTGGAAGATATCTACCCCTCCGAGGCATCTGGAACCGAGAGCAAGAA	128
Qy	440	GCTGAACCGGCGAGTACACCCCTCTGAGCATCTGACCATCTGACCAACGAGCGGGGCGGG	499
Db	129	GCTGAACCGGCGAGTACACCCCTCTGAGCATCTGACCATCTGACCAACGAGCGGGGCGGG	188

QY	500	CAGGCTCTCCAAATATAAGCCATCTGAACCTACTGGAACCTGACCTGAGGTCACGTCGGT	559
Db	189	CAGGCTCTCCAAATATAAGCCATCTGAACCTACTGGAACCTGACCTGAGGTCACGTCGGT	249
QY	560	GGTTACGGGTTTTAAACAACCTGCGGACAGATTTAAAGACATTTCTGCTGATTTGACATG	619
Db	249	GGTTACGGGTTTTAAACAACCTGCGGACAGATTTAAAGACATTTCTGCTGATTTGACATG	308
QY	620	CCGCAATTTTACCTGCTTATAGATACAGCCGGATTAAGTGTGCAAGAACAACCTTTCTTGT	679
Db	309	CCGCAATTTTACCTGCTTATAGATACAGCCGGATTAAGTGTGCAAGAACAACCTTTCTTGT	368
QY	680	GCTGGCATTAAAGTCCCTCACTCCACATTTTGCCAGAAGGCAACAATCCGGGAATCCTG	739
Db	369	GCTGGCATTAAAGTCCCTCACTCCACATTTTGCCAGAAGGCAACAATCCGGGAATCCTG	428
QY	740	GGGCCAAGAAGCAACCGCAGGGAACCAACCGGTGGTGGAGTCTTCTGCTGGGCCAGAC	799
Db	429	GGGCCAAGAAGCAACCGCAGGGAACCAACCGGTGGTGGAGTCTTCTGCTGGGCCAGAC	488
QY	800	ACCCCCAGAGGACAACCCACCGGACCTTTCAGATATGCTGAAATTTTCAGAGTGAGAAGCA	859
Db	489	ACCCCCAGAGGACAACCCACCGGACCTTTCAGATATGCTGAAATTTTCAGAGTGAGAAGCA	548
QY	860	CCAGACATTTTATGTGGAACCTACAGACACACTTTCTTCAACTTGTCTGGAAGGAAGT	919
Db	549	CCAGACATTTTATGTGGAACCTACAGACACACTTTCTTCAACTTGTCTGGAAGGAAGT	608
QY	920	GCTCTTTCTCAGTGGGTAACTACTCTCCGCCAGACACTGAGTTCTTTCAGGGCGA	979
Db	609	GCTCTTTCTCAGTGGGTAACTACTCTCCGCCAGACACTGAGTTCTTTCAGGGCGA	668
QY	980	TGACGATGTTTTTGTGAACCCCATCATCTCTCAATTAATTTTGAATTTTATCCAAG	1037
Db	669	TGACGATGTTTTTGTGAACCCCATCATCTCTCAATTTTGAATTTTATCCAAG	728
QY	1038	ACCAAGGCC-AAAGATCTCTCATAGG-TGATGTGATCCCAATGCT-GGACCTCATCGG	1094
Db	729	ACCAAGGCCAAAGATTTCTTCAAAAGTTGAATGATCCCAATGCTGGGACCTCATCGG	788
QY	1095	GATAAGAGCTGAAG-TACTACATCCCGAAGTGTGT	1131
Db	789	GATAAGAGCTGAAGTTACTATACATCCCANAAAGTTGT	826
RESULT 3			
LOCUS	BX370083	876 bp	mRNA
DEFINITION	BX370083 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DI074YK14 5-PRIME, mRNA sequence.		
ACCESSION	BX370083		
VERSION	BX370083.1	GI:30449879	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 876)		
JOURNAL	Li.W.B., Gruber,C., Jessee,J. and Polayes,D.		
COMMENT	Full-length cDNA libraries and normalization		
	Unpublished		
	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	BP 191 91006 Evry cedex - France		
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr		
	Library was constructed by Life Technologies, a division of		
	Invitrogen. This sequence belongs to sequence cluster 1713.f For		
	more information about this cluster, see		
	http://www.genoscope.cns.fr/		
	cgi-bin/cluster.cgi?seq=CS0BAG0362C11_CS03443.1&cluster=1713.f.		
	Contact : Feng Liang Email : fliang@lifetech.com URL :		
	http://fulllength.invitrogen.com/ Invitrogen Corporation 1600		


```

QY 976 GCGATGACGAGTGTCTTTGTGAACACCCATCACATCCIGAAATTAATTAATGTTATCCA 1035
D 490 GCGATGACGAGTGTCTTTGTGAACACCCATCACATCCIGAAATTAATTAATGTTATCCA 549
QY 1036 AGACCAAGCCAAAGATCTCTTCATAGGTGATGTGATCCACATCTGACCTCATCGGG 1095
D 550 AGACCAAGCCAAAGATCTCTTCATAGGTGATGTGATCCACATCTGACCTCATCGGG 609
QY 1096 ATRAGAAGCTGAAGTACTACATCCCA--GAAGTGTGTTTACTCTGGCTCTACCCACCCCTA 1153
D 610 ATRAGAAGCTGAAGTACTACATCCCAAGGAAGTGTGTTTACTCTGGCTCTACCCACCCCTA 669
QY 1154 TGCAGGGGGAGGGGGTTCCTC 1175
D 670 TGCAGGGGGAGGGGGTTCCTC 691

RESULT 7
LOCUS BM557603 1053 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6563386 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5739980
S.: mRNA sequence.
ACCESSION BM557603
VERSION BM557603.1 GI:18799721
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1053)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12754 row: c column: 21
High quality sequence stop: 547.
FEATURES
source
1..1053
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5739980"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_88"
/notes="Organ: small intestine; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
oligo-3T primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC library."
BASE COUNT 288 a 240 c 219 g 306 t
ORIGIN
Query Match 38.0%; Score 648.6; DB 12; Length 1053;
Best Local Similarity 94.0%; Pred. No. 1.8e-138;
Matches 707; Conservative 8; Mismatches 27; Indels 10; Gaps 4;
QY 870 CTTATGTGGAACACTACAGACACTTCTTCAACTTGTCTCTGAAGGAAGTGTGTTTCTC 929
D 1 CTTATGTGGAACACTACAGACACTTCTTCAACTTGTCTCTGAAGGAAGTGTGTTTCTC 60
QY 930 AGGTGGGTAACTACTCTGCCACACTGAGTTTCTTCAAGGGCGGATGACGATGT 989
D 61 AGGTGGGTAACTACTCTGCCACACTGAGTTTCTTCAAGGGCGGATGACGATGT 120

```

```

QY 990 TTTGTGAACACCCATCACATCTCTGAATTAATTAATGTTTATCCAAAGACCAAGCCAAA 1049
D 121 TTTGTGAACACCCATCACATCTCTGAATTAATTAATGTTTATCCAAAGACCAAGCCAAA 180
QY 1050 GATCTCTTCATAGGTGATGTGATCCACATGCTGGACCTCATCGGATAGAAGTGAAG 1109
D 181 GATCTCTTCATAGGTGATGTGATCCACATGCTGGACCTCATCGGATAGAAGTGAAG 240
QY 1110 TACTACATCCCAAGAGTGTGTTTACTCTGGCTCTACCCACCTATACAGGGGAGGGGG 1169
D 241 TACTACATCCCAAGAGTGTGTTTACTCTGGCTCTACCCACCTATACAGGGGAGGGGG 300
QY 1170 TTCTCTTACTCCGGCCACCTGGCCCTGAGCTCTACATATCACTGACGAGTCCATCTC 1229
D 301 TTCTCTTACTCCGGCCACCTGGCCCTGAGCTCTACATATCACTGACGAGTCCATCTC 360
QY 1230 TACCCCATTTGATGACGTTTATCTGGAATGTGCTTCAGAACTCGGCTCGTTCAGAG 1289
D 361 TACCCCATTTGATGACGTTTATCTGGAATGTGCTTCAGAACTCGGCTCGTTCAGAG 420
QY 1290 AAACACAAAGGCTTCAGACATTTGATATCGAGAGAGAAAACAAAATACATCTCTCC 1349
D 421 AAACACAAAGGCTTCAGACATTTGATATCGAGAGAGAAAACAAAATACATCTCTCC 480
QY 1350 TATGATAGTCTGATGTTAGTACATAGTAGAAAAACCTCAAGAGATGATGATTTGGTCT 1409
D 481 TATGATAGTCTGATGTTAGTACATAGTAGAAAAACCTCAAGAGATGATGATTTGGTCT 540
QY 1410 CAGTTGAGAGTGTCTCATTTAAATGCTAAATAGATACAACTCAATTKGSATWGRA 1469
D 541 CAGTTGAGAGTGTCTCATTTAAATGCTAAATAGATACAACTCAATTTGATAGAAA 600
QY 1470 GGG-GTWTGTTGATGAGGCTGTTGGGTCTGCTCAATAGAGTAATTTCTATTTAAAN 1528
D 601 GGGGTGATTTGATGATGTTGCTGCTGCTCAATAGAGTAATTTCTATTTAAAN 660
QY 1529 C---ATGAATTTGCTTTATGAGTGTATACCATTT--ANGGCTCTAANCCCTT---CAT 1579
D 661 CCCTGAAATTTGGCTTTTATGAGGATACCCCTTTGAGGGGCTCTAAGCCCTTCAATT 720
QY 1580 TTGNACTCAGCTGAAGAGGAAAGCGGAGA 1611
D 721 TGGTACTCCCTGGAAGAGGAAAGCGGAAA 752

RESULT 8
LOCUS B1549904 725 bp mRNA linear EST 05-SEP-2001
DEFINITION mRNA sequence.
ACCESSION B1549904
VERSION B1549904.1 GI:15437216
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 725)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11692 row: d column: 14
High quality sequence stop: 680.

```



```

QY 309 TATTTTATTATGAGTCTCCAAAGCAGTAGCCACAGAAAAAATGGAAGAGGGGAAGTA 368
DB 301 IATTTTATTATGAGTCTCCAAAGCAGTAGCCACAGAAAAAATGGAAGAGGGGAAGTA 360
QY 369 ATAATACCCAAAGAGAGTCTTGGAGATATCTACCCCTCCCGAGGCATCTGGAAACCGA 428
DB 361 ATAATACCCAAAGAGAGTCTTGGAGATATCTACCCCTCCCGAGGCATCTGGAAACCGA 420
QY 429 GAGCAAGAGAGCTGAAACCGGAGTAAACCCCATCTGAGCATGCTGACCAACAGAGG 488
DB 421 GAGCAAGAGAGCTGAAACCGGAGTAAACCCCATCTGAGCATGCTGACCAACAGAGG 480
QY 489 GGGGAGCGGCGAGGCTCTCCATATAAGCATCTGAACTACTGCAACCTGACCTGAGG 548
DB 481 GGGGAGCGGCGAGGCTCTCCATATAAGCATCTGAACTACTGCAACCTGACCTGAGG 540
QY 549 GTCACGTGGTGGTTACAGGTTTAAACACTTCCCGGACAGATTTAAAGACTTT 602
DB 541 GTCCCGTGGTGGTACAGGTTTAAACACTTCCCGGACAGATTTAAAGAGTTT 594

RESULT 10
BG741357
LOCUS 602631909P1 NCI_CGAP_Skn3 659 bp mRNA linear EST 15-MAY-2001
DEFINITION 602631909P1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4776866 5',
mRNA sequence.
ACCESSION BG741357
VERSION BG741357.1 GI:14052010
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 659)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Prepared by: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10630 row: b column: 03
High quality sequence stop: 659.
Location/Qualifiers
FEATURES
source
1..659
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4776866"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Skn3"
/Note="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: oligo dh.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 173 a 168 c 155 g 163 t
ORIGIN
Query Match 33.1%; Score 565; DB 10; Length 659;
Best Local Similarity 96.1%; Pred. No. 3e-119;
Matches 635; Conservative 0; Mismatches 15; Indels 11; Gaps 5;
QY 586 ACAGATTAAAGACTTCTGCTGATTGTGAGATGCCCAATATTCTACTGCTTAATGATC 645
DB 1 ACAGATTAAAGACTTCTGCTGATTGTGAGATGCCCAATATTCTACTGCTTAATGATC 60
QY 646 AGCCGGATAAGTGTGCAAGAAACCTTCTTGTGTGGCCATTAAAGTCCCTCACTCCAC 705
|||||

```

```

DB 61 AGCCGGATAAGTGTGCAAGAAACCTTTCTTG-TGCTGGGATTAAAGTCCCTCACTCCAC 119
QY 706 ATTTTGCAGAGGCAAGCAATCCGGGAATCTCTGGGGCCCAAGAAACCAACGCGAGGAACC 765
DB 120 ATTTTGCAGAGGCAAGCAATCCGGGAATCTCTGGGGCCCAAGAAACCAACGCGAGGAACC 179
QY 766 AAACGGTGGTGGGAGTCTTCTGCTGGCCAGACACCCCGAGAGGACCAACCCCGAGCC 825
DB 180 AAACGGTGGTGGGAGTCTTCTGCTGGCCAGACACCCCGAGAGGACCAACCCCGAGCC 239
QY 826 TTTTCAGATATGCTGAAATTTGAGAGTGAGAACCAACCAAGACATTTCTTATGTGGAATACA 885
DB 240 TTTTCAGATATGCTGAAATTTGAGAGTGAGAACCAACCAAGACATTTCTTATGTGGAATACA 299
QY 886 GAGACACTTTCTCAACTTGTCTGCTGAAGGAAGTGTGTTTC-----TCAGGTGGGTAAG 940
DB 300 GAGACACTTTCTCAACTTGTCTGCTGAAGGAAGTGTGTTTCAGGTGGGTAAG 359
QY 941 TACTTCC-TGCCAGACACACTGAGTTTCTTTTC---AAGGGCGATGACGATGTTTTTGTGA 996
DB 360 TACTTCCGTCCCGACAGACGATGAGTTGTTCTCAAGGGCGGATGACGATGTTTTGTGA 419
QY 997 ACACCATACATCCTGAATTACTTTAGTAGTTTATCCAGAGACCAAGCCAAAGATCTCT 1056
DB 420 ACACCATACATCCTGAATTACTTTAGTAGTTTATCCAGAGACCAAGCCAAAGATCTCT 479
QY 1057 TCATAGGTGATGTCATCCACAATGCTCGGACCTCATCGGATAAGAAAGCTGAATCTACTACA 1116
DB 480 TCATAGGTGATGTCATCCACAATGCTCGGACCTCATCGGATAAGAAAGCTGAATCTACTACA 539
QY 1117 TCCAGAGATTGTTTACTTCTGGCCTCTACCCACCTATGCAAGGGGAGGGGGTTCCTCT 1176
DB 540 TCCAGAGATTGTTTACTTCTGGCCTCTA-CCACCTATGCAAGGGGAGGGGGTTCCTCT 598
QY 1177 ACTCGGGCACCTGGCCCTGAGGCTGACCATATCAGTACCAGGTCCATCTCTACCCCA 1236
DB 599 ACTCGGGCACCTGGCCCTGAGGCTGACCATATCAGTACCAGGTCCATCTCTACCCCA 658
QY 1237 T 1237
DB 659 T 659

RESULT 11
BG721682
LOCUS 602695830P1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:482158 5',
DEFINITION mRNA sequence.
ACCESSION BG721682
VERSION BG721682.1 GI:14000869
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 632)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Prepared by: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10745 row: k column: 07
High quality sequence stop: 625.
Location/Qualifiers
FEATURES
source
1..632

```



```

Qy 761 GAACAAACGGTGGTGGAGTCTTCTGCTGGCCAGACACCCCGAGGAGCAACACCC 820
Db 595 GAACAAACAGTGGTGGAGTCTTCTGCTGGCCAGACACCCCGAGGAGCAACACCC 654
Qy 821 CGACCTTTCAGAI 833
Db 655 TGACCTTTCRAAT 667

RESULT 13
AV724042
LOCUS AV724042 HTB Homo sapiens cDNA clone HTBAXG01 5', mRNA linear EST 16-OCT-2000
DEFINITION AV724042 HTB Homo sapiens cDNA clone HTBAXG01 5', mRNA sequence.
ACCESSION AV724042
VERSION AV724042.1 GI:10828039
KEYWORDS EST.
SOURCE EST.
ORGANISM Homo sapiens (human)
Unpublished
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex. 45)
Fax: 86-21-50801922
Email: hanzq@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1..701
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HTBAXG01"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="HTB"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 183 a 159 c 177 g 177 t 5 others
ORIGIN
Query Match 28.6%; Score 488.2; DB 9; Length 701;
Best Local Similarity 93.2%; Pred. No. 1.5e-101;
Matches 564; Conservative 0; Mismatches 32; Indels 9; Gaps 5;

Qy 489 GGGAGGGGGCGGCTCTCCATATAGCCATCTGAAGTACTGCGAAGCTGACCTGAGG 548
Db 103 GGAGAACAGAGAGCTCTCCATATAGCCATCTGAAGTACTGCGAAGCTGACCTGAGG 162
Qy 549 GTCACGTGGTGGTTACGGGTTTAAACAACCTCCGCGACAGATTTAAAGACTTTCTGCTG 608
Db 163 GTCACGTGGTGGTTACGGGTTTAAACAACCTCCGCGACAGATTTAAAGACTTTCTGCTG 222
Qy 609 TATTGAGATGCCGCATTTATCTACTGCTTATAGATCAGCGGATAGTGTGCAAGAAA 668
Db 223 TATTGAGATGCCGCATTTATCTACTGCTTATAGATCAGCGGATAGTGTGCAAGAAA 282
Qy 669 CCTTCTTGTGCTGGCGATTAACTCCCTCACTCCACATTTGCCAGAGGCAAGCAATC 728
Db 283 CCTTCTTGTGCTGGCGATTAACTCCCTCACTCCACATTTGCCAGAGGCAAGCAATC 342
Qy 729 CGGGAATCCTGGGCGCCAGAAAGCAACGCGAGGGAACCAACGGTGGTGGAGTCTTCCTG 788
Db 343 CGGGAATCCTGGGCGCCAGAAAGCAACGCGAGGGAACCAACGGTGGTGGAGTCTTCCTG 402

```

```

Qy 789 CTGGGCGAGACACCCCGAGAGAACACCCCGAGCTTTTCAGATATGCTGAAATTTGAG 848
Db 403 CTGGGCGAGACACCCCGAGAGAACACCCCGAGCTTTTCAGATATGCTGAAATTTGAG 462
Qy 849 AGTCAGAGCACCACCAAGACATTTATGTGCACTCAGAGACACTTTCTTCAACTTGTCT 908
Db 463 AGTCAGAGCACCACCAAGACATTTATGTGCACTCAGAGACACTTTCTTCAACTTGTCT 522
Qy 909 CTGAAGGAAGTGTGTTTCTCAGTGGTGAAGTACTTCTGCGCCAGACACTGAGTTTGT 968
Db 523 CTGAAGGAAGTGTGTTTCTCAGTGGTGAAGTACTTCTGCGCCAGACACTGAGTTTGT 582
Qy 969 TTCAAGGGCGATGACGATGTTTTTGTGTAACCCATCAGATCTCTGAATTTACTGATAG 1028
Db 583 TTCAAGGGCGATGACGATGTTTTTGTGTA--CACCTTACATCCGGATTTACTTGT-ATAGG 639
Qy 1029 TTATCCAAAGACCAAGCCAAAGATCTCTCATAGGTGATGATGATGATGATGATGATG 1088
Db 640 TTATCCAAAGACCAAGCCAAAGATCTCTCATAGGTGATGATGATGATGATGATGATG 693
Qy 1089 CATCG 1093
Db 694 ATTGG 698

RESULT 14
BF001654/c
LOCUS BF001654
DEFINITION 7991C06.x1 NCI-CGAP_col6 Homo sapiens cDNA clone IMAGE:3313834 f'
sequence.
ACCESSION BF001654
VERSION BF001654.1 GI:10701929
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 608)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -400p from Gibco
High quality sequence stop: 465.
FEATURES
source
1..608
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3313834"
/tissue_type="Colon tumor, RER+"
/lab_host="DH10B"
/clone_lib="NCI-CGAP_col6"
/notes="Organ: colon; Vector: pT73D-pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP Col10 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351)."

```

```
BASE COUNT      181 a  128 c  121 g  178 t
ORIGIN
Query Match      27.2%; Score 463.6; DB 10; Length 608;
Best Local Similarity 91.6%; Pred. No. 6.5e-96;
Matches 522; Conservative 8; Mismatches 33; Indels 7; Gaps 4;
QY 1068 GTATCCACAATGCTGGACCTCATCGGCATAGCAAGCTGAAGTACTACATCCAGAGIT 1127
DB 608 GTATCCACAATGCTGGACCTCATCGGTGAAGTAGCTGAAGTACTACATCTCAGTAGIT 549
QY 1128 GTTACTCGGCTCTACCCACCTATGCGAGGGAGGGGCTCTCTACTCCGGCCAC 1187
DB 548 GTTACTCGGCTCTACACACATATGCGAGGTGTAGGAGGTTCCTTACTCCGGCCAC 489
QY 1188 CTGGCCCTGAGGCTGTACCATATATCATGACAGGTGCATCTTACCCCATTTGAAGCATT 1247
DB 488 CTGGCCCTGAGGCTGTACCATATATCATGACAGGTGCATCTTACCCCATTTGAAGCATT 429
QY 1248 TATACTGGAATGTCCCTTCAGAACTCGGCCCTGTTCCAGAGAAACACAAGGCTTCAGG 1307
DB 428 TATACTGGAATGTCCCTTCAGAACTCGGCCCTGTTCCAGAGAAACACAAGGCTTCAGG 369
QY 1308 ACATTTGATATCGAGGAGAAAACAAAATAAATACATCTGCTCTATGTAGATCTGATGTTA 1367
DB 368 ACATTTGATATCGAGGAGAAAACAAAATAAATACATCTGCTCTATGTAGATCTGATGTTA 309
QY 1368 GTACATAGTAGAAACCTCAAGAGATGATTGATATTTGGTCTCAGTTGCGAGAGTGCAT 1427
DB 308 GTACATAGTAGAAACCTCAAGAGATGATTGATATTTGGTCTCAGTTGCGAGAGTGCAT 249
QY 1428 TTTAAATGCTAAATATAGATACAACTCAATTTGSGATGAAAGGGTWTTTTGRATGGY 1487
DB 248 TTTAAATGCTAAATATAGATACAACTCAATTTGSGATGAAAGGGTWTTTTGRATGGY 189
QY 1488 CCCATGTTGGGCTCATATAGATGATTTCTATTNAANC--ATGAAATTCGCTTTAT 1545
DB 188 CCCATGTTGGTCTCACA-TAGAGIAATTC-ATATTAACCATGAATATGCCITAT 129
QY 1546 GAGTGATACCCATTT-ANGGCTCTAANCCTT---CATTTGNACTACGTGAAGAAGGA 1601
DB 128 GAGTGATACCCATTTGAGGCGCTCAACCCCTTCAATTTGTTACTCAGTGAAG-AGGGA 70
QY 1602 AAGCGGGAGAGGTAATTTTATGCTGA 1631
DB 69 AAGCGGAGATGCTAATTTCTTTATGGA 40

RESULT 15
AW958548
LOCUS      AW958548          593 bp  mRNA  linear  EST 01-JUN-2000
DEFINITION EST370618 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW958548
VERSION    AW958548.1  GI:8148232
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
```

```
Plate: 121
Seq primer: Reverse.
Location/Qualifiers
1..593
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGE"
/note="Vector: pBluescriptSKm"
BASE COUNT      127 a  112 c  114 g  109 t  131 others
ORIGIN
Query Match      26.9%; Score 458.8; DB 10; Length 593;
Best Local Similarity 99.6%; Pred. No. 8.4e-95;
Matches 460; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 708 TTGGCAGAAAGCAAGCAATCCGGGAATCCTGGGGCCAAAGAAAGCAACCGAGGAACCAA 767
DB 1 TTGGCAGAAAGCAAGCAATCCGGGAATCCTGGGGCCAAAGAAAGCAACCGAGGAACCAA 60
QY 768 ACGGTGGTGGGAGTCTTCTGCTGGGCCAGACACCCCCAGAGAGACACACCCCGACCTT 827
DB 61 ACGGTGGTGGGAGTCTTCTGCTGGGCCAGACACCCCCAGAGAGACACACCCCGACCTT 120
QY 828 TCAGATATGCTGAAATTTGAGAGTGAGAAGCACCACCAAGACATCTTATGTGGAACACAGA 887
DB 121 TCAGATATGCTGAAATTTGAGAGTGAGAAGCACCACCAAGACATCTTATGTGGAACACAGA 180
QY 888 GACACTTCTTCAACTTGTCTCTGAAGGAAGTGCTGTTTCTCAGGTGGGTAAGTACTTCC 947
DB 181 GACACTTCTTCAACTTGTCTCTGAAGGAAGTGCTGTTTCTCAGGTGGGTAAGTACTTCC 240
QY 948 TGCCCGACACACTGAGTTTCTTCAAGGGCGATGACGATGTTTCTGGAACACCCCATCAC 1007
DB 241 TGCCCGACACACTGAGTTTCTTCAAGGGCGATGACGATGTTTCTGGAACACCCCATCAC 300
QY 1008 ATCCTGAATTACTTGAATAGTTTATCCAAAGACCAAGCCAAAGATCTCTTCATAGTGAT 1067
DB 301 ATCCTGAATTACTTGAATAGTTTATCCAAAGACCAAGCCAAAGATCTCTTCATAGTGAT 360
QY 1068 GTGATCCACAATGCTGGACCTCATCGGGATAGAAAGCTGAAGTACTACATCCAGAGATT 1127
DB 361 GTGATCCACAATGCTGGACCTCATCGGGATAGAAAGCTGAAGTACTACATCCAGAGATT 420
QY 1128 GTTACTCTGGCCTCTACCCACCTATGCGAGGGGAGGGGG 1169
DB 421 GTTACTCTGGCCTCTACCCACCTATGCGAGGGGAGGGGG 462
```

Search completed: October 20, 2003, 19:37:17
Job time : 3863.36 secs


```

RESULT 4
US-09-482-180A-1
; Sequence 1, Application US/09482180A
; Patent No. 6361985
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Yamamoto, Gayle
; APPLICANT: Gao, Zeren
; APPLICANT: Whitmore, Stephen
; APPLICANT: Jasper, Stephen
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOG, ZN5SP6
; FILE REFERENCE: 98-80
; CURRENT APPLICATION NUMBER: US/09/482.180A
; CURRENT FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/115,721
; PRIOR FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1420
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (135)...(1271)
US-09-482-180A-1

```


Qy	1088	TCATCGGGATAGAAGCIGAACTACTACATCCAGAGAGTTGTTTACTCTCIGGCCTCTACCC	1147
Db	928	GCCTCCGAAGCCCGAGGAGCCCTTCTATGTGCCGAAGACCTTCTTTTGAAGGGGACTATCC	987
Qy	1148	ACCCTATCGAGGGGAGGGGGTTCCTCTACTCCGGCCACCTGSCCCTCAGGCTGTACCA	1207
Db	988	AGCCTATCGAGTGAGAGTGGCTATGTAACTCTCAGGACGGCTGGCACCTTGGCTGTGCA	1047
Qy	1208	TATCACTGACCAAGCTCCATCTCTACCCCAITGATGACGTTTATCTAGGAATGTGCCCTCA	1267
Db	1048	GGCGGCAGCTCGGTGTGCACCCCTTCCCCCTTTGATGATGTCTACACTGSCCTTCTGCTTCG	1107
Qy	1268	GAAACTCGGCTCGCTTCCAGAGAAACACAAGGCTTCAGGACAATTGATATCGAGGAGAA	1327
Db	1108	TGCCCTGGCTTAGCACCCCGTGCCCATCCAGGCTTCCCTACA---GCCTGGCCAGCAGA	1164
Qy	1328	AAACAAAAATACATCTGCTCCTATGTAGATCTCATGTGTAGTACATAGTAGNAACCTCA	1387
Db	1165	ACGTACCAGGAGACCCCTGGCGCGTGGAGGGCGTCTCTTGGTGCATCCAGTCAGCCCTCA	1224
Qy	1388	AGAGATGATTGATATTTGGTCTCAGITTCAGAGAGTGCCTCATTTAAATATGCTAA	1439
Db	1325	GGACACCATTTGGCTCTGGAGACATCTGTGGGTCCTCCAGAGCTCCAGCTGCTGA	1276

RESULT 7

```

US-09-996-243-208
: Sequence 208, Application US/09996243
: Patent No. 6478825
: GENERAL INFORMATION:
: APPLICANT: Ashkenazi, Avi J.
: APPLICANT: Baker, Kevin P.
: APPLICANT: Botstein, David
: APPLICANT: Desnoyers, Luc
: APPLICANT: Eaton, Dan L.
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Fong, Sherman
: APPLICANT: Gerber, Hanspeter
: APPLICANT: Gerritsen, Mary E.
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, J. Christopher
: APPLICANT: Gurney, Austin L.
: APPLICANT: Kijavini, Ivar J.
: APPLICANT: Napier, Mary A.
: APPLICANT: Pan, James
: APPLICANT: Paoni, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: TITLE OF INVENTION: Acids Encoding the Same
: FILE REFERENCE: P2730PlC13
: CURRENT APPLICATION NUMBER: US/09/996,243
: CURRENT FILING DATE: 2001-11-14
: PRIOR APPLICATION NUMBER: 60/049787
: PRIOR FILING DATE: 1997-06-16
: PRIOR APPLICATION NUMBER: 60/062250
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/065186
: PRIOR FILING DATE: 1997-11-12
: PRIOR APPLICATION NUMBER: 60/065311
: PRIOR FILING DATE: 1997-11-13
: PRIOR APPLICATION NUMBER: 60/066770
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 60/075945
: PRIOR FILING DATE: 1998-02-25
: PRIOR APPLICATION NUMBER: 60/078910
: PRIOR FILING DATE: 1998-03-20

```


; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02

; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 3.1%; Score 53.6; DB 4: Length 2095;
Best Local Similarity 45.7%; Pred. No. 0.00031;
Matches 308; Conservative 0; Mismatches 354; Indels 12; Gaps 3;
QY 591 TTTAAAGACTTTCTGCTGCTATTTGAGATGCCGCAATTAITCACTGCTATAGATCAGCG 650
DB 475 TTTATGAGTATGAGCCGATTTACAGCAACACTTTCACTTCACACTTCGAGACATCA 534
QY 651 GATAAGTGTGCAAGAAACCTTTCTTTGTTGGCGATTAAAGTCCCTCACTCCCAATTTT 710
DB 535 AACTGCTCTCATCAAAATCCATTTCTGTGTCATTTCTGGTGACCTCCACCCCTTCAGATGTG 594
QY 711 GCCAGAAGCGAAGCAATCCGGGAATCCTGGGGCCCAAGAAACCAAGCAGGAGCAACG 770
DB 595 AAAGCCAGGCGAGCCATTTAGAGTTACTTTGGGTGAAAAAAGTCTTGGTGGGATATGAG 654
QY 771 GTGGTGGAGTCTTCTGCTGGCGAGACACCCCGAGAGACACCCCGACCTTTCA 830
DB 655 GTTCTTACATTTTCTTATTAGGCCAAGAGCGCTGAAAGGAGAGACAATG---ITGGCA 711
QY 831 GATATGCTGAATTTGAGAGTGAGAACCAACACATCTTTATGTGGAACCTACAGAGAC 890
DB 712 TTGCTTTAGAGGATGAACACCTTTTATGGTGACATAATCCGACAGATTTTITAGAC 771
QY 891 ACTTTCTCAACTGCTCTCTGAAGCACTGCTGTTTCTCAGGTGGTAAAGTACTTCTGTC 950
DB 772 ACATATATACTGACCTGCTGCTGAAACCACTTATGGCATTCAGGTGGGTAACTGAGTTTGC 831
QY 951 CCAGACACTGATGTTGTTTCAAGGGCGATGACGATGTTTGTGAACACCCCATCATC 1010
DB 832 CCCAATGCCAAGTACGTAATGAAGACACACACTGATGTTTTCATCAATACTGSCAATTA 891
QY 1011 CTGAATTTAGTATGTTTATCCAGACCAAGCAACCAAGATCTTTCATAGGTGATGTG 1070
DB 892 GTGAA-----GTATCTTTTAAACCTAAACCACTCAGAGAGTTTTTCACAGGTATCCCT 945
QY 1071 ATCCCAATGCTGGACCTCATCGGATGAAGAGTGAAGTACTATATCCCAAGAGTGT 1130
DB 946 CTAATTGATAATTATTCATATAGAGGATTTTACCAAAAACCCATATTCCTTACCAGAG 1005
QY 1131 TACTCTGGC---CTCTACCCACCTATGCGAGGGGAGGGGGTTCCTCTACTCCGGCCAC 1187
DB 1006 TATCCTTTCAAGGTGTTCCCTCCATACCTGTCAGTGGGTGGGTATATAATGTCAGAGAT 1065
QY 1188 CTGGCCCTGAGGCTGTACCATATCACTACAGGTCCTCATCTACCCATGATGAGGTT 1247
DB 1066 TTGGTGCCAAGATCTATGAATATGATGGTCACGTAAACCCCAACCAAGTTTGAAGATGT 1125
QY 1248 TATACTGGAATGTG 1261
DB 1126 TATGTCGGGATCTG 1139

RESULT 8

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS

QY 61 AGCGGAGACGTCCTCCCGCCGACACCCCGCGCGCGCGTCCGCGCGCGCGCGCATGG 120
DB 61 AGCGGAGACGTCCTCCCGCCGACACCGCGCGCGCGCGTCCGCGCGCGCGCGCATGG 120
QY 121 AGCGTGAGCTGCGCGCGTCCCGCGCTGAGCGCGCGCGGAGCGCGCGCGCGCATGG 190
DB 121 AGCGTGAGCTGCGCGCGTCCCGCGCTGAGCGCGCGCGGAGCGCGCGCGCGCATGG 180
QY 181 CCGCGATCTCCCGCCCTTGCCCGCCCGCCCGCGCGAGCTGAGCTGCTCCCGGACAGATAT 240
DB 181 CCGCGATCTCCCGCCCTTGCCCGCCCGCCCGCGCGAGCTGAGCTGCTCCCGGACAGATAT 240
QY 241 GAGAAATGAGTGTGGACGTCGGAAGATAAGTGTGTTGGGTATCTCTGATGATGGCAATG 300
DB 241 GAGAAATGAGTGTGGACGTCGGAAGATAAGTGTGTTGGGTATCTCTGATGATGGCAATG 300
QY 301 TCTTCATTTATTTATGGAAGTCTCCAAAGCAGTACCCACCCATCTCTGAGCACTGACCA 360
DB 301 TCTTCATTTATTTATGGAAGTCTCCAAAGCAGTACCCACCCATCTCTGAGCACTGACCA 360
QY 361 GGAAGTAAATATACCAAGAGAGCTCTGGAAGATATCTACCCCTCCCGAGGCATCT 420
DB 361 GGAAGTAAATATACCAAGAGAGCTCTGGAAGATATCTACCCCTCCCGAGGCATCT 420
QY 421 GGAACCGAGAGCAAGAGAGCTGAACCGCGCAGTACAAACCCATCTCTGAGCACTGACCA 480
DB 421 GGAACCGAGAGCAAGAGAGCTGAACCGCGCAGTACAAACCCATCTCTGAGCACTGACCA 480
QY 481 ACCAGACGGGGAGCGGGGAGGCTCTCCAAATTAAGCCATCTGAACTACTGCGAAGCTG 540
DB 481 ACCAGACGGGGAGCGGGGAGGCTCTCCAAATTAAGCCATCTGAACTACTGCGAAGCTG 540
QY 541 ACCTGAGGTCACGTCGGTGGTTACGGGTTTAAACAACTGCGCGACAGATTTAAAGACT 600
DB 541 ACCTGAGGTCACGTCGGTGGTTACGGGTTTAAACAACTGCGCGACAGATTTAAAGACT 600
QY 601 TCTGCTGATTTGAGATGCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 660
DB 601 TCTGCTGATTTGAGATGCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 660
QY 661 CAAGAAGAACCTTTCTTCTGCTGGCGATTAAGTCCCTCACTCCACATTTTCCAGAGGC 720
DB 661 CAAGAAGAACCTTTCTTCTGCTGGCGATTAAGTCCCTCACTCCACATTTTCCAGAGGC 720
QY 721 AAGCAATCCGGGAATCTGGGGCAAGAAAGCAACCGAGGAAACCGGAGTGGTCCGAG 780
DB 721 AAGCAATCCGGGAATCTGGGGCAAGAAAGCAACCGAGGAAACCGGAGTGGTCCGAG 780
QY 781 TCTTCTGCTGGCGCAGACACCCCGAGAGGACACACCCCGACCTTTCAGATATCTGA 840
DB 781 TCTTCTGCTGGCGCAGACACCCCGAGAGGACACACCCCGACCTTTCAGATATCTGA 840
QY 841 AATTGAGGTGAGAAGCAACCAAGACATCTTATGTGGAACCTACAGAGACACTTCTCA 900
DB 841 AATTGAGGTGAGAAGCAACCAAGACATCTTATGTGGAACCTACAGAGACACTTCTCA 900
QY 901 ACTTGTCTGAGAAGAGTCTGTTTCTCAGGTGGTAACTACTTCTGCGGACAGACTG 960
DB 901 ACTTGTCTGAGAAGAGTCTGTTTCTCAGGTGGTAACTACTTCTGCGGACAGACTG 960
QY 961 AGTTGTTTTCAAGGGGATGAGATGTTTCTGAAACCCCATCATCTCAATTAAT 1020
DB 961 AGTTGTTTTCAAGGGGATGAGATGTTTCTGAAACCCCATCATCTCAATTAAT 1020
QY 1021 TGAATAGTTTATCAAGACCAAGCCAAAGATCTCTTCATAGGTGATGATCCCAATG 1080
DB 1021 TGAATAGTTTATCAAGACCAAGCCAAAGATCTCTTCATAGGTGATGATCCCAATG 1080
QY 1081 CTGAGCTCTACCGGATGAAGAGCTGAAGTACTACATCCAGAGTGTGTTTACTCTGGCC 1140
DB 1081 CTGAGCTCTACCGGATGAAGAGCTGAAGTACTACATCCAGAGTGTGTTTACTCTGGCC 1140

QY 1141 TCTACCCACCTATGTCAGGGGAGGGGGTTCCTCTACTCGGCGACCTGGCCCTGAGGC 1200
DB 1141 TCTACCCACCTATGTCAGGGGAGGGGGTTCCTCTACTCGGCGACCTGGCCCTGAGGC 1200
QY 1201 TGTACCATATCACTGACGAGCTCCATCTCTACCCCATTTGATGAGCTTTATCTAGGAATG 1260
DB 1201 TGTACCATATCACTGACGAGCTCCATCTCTACCCCATTTGATGAGCTTTATCTAGGAATG 1260
QY 1261 GCCTTCAGAAACTCGGCTCTGTCAGAGAAACAAGGCTTCAGGACATTTGATATCG 1320
DB 1261 GCCTTCAGAAACTCGGCTCTGTCAGAGAAACAAGGCTTCAGGACATTTGATATCG 1320
QY 1321 AGGAGAAAACAATAAATCAATCTGCTCTATGTAGATCTGATCTTAGTACATAGTAGAA 1380
DB 1321 AGGAGAAAACAATAAATCAATCTGCTCTATGTAGATCTGATCTTAGTACATAGTAGAA 1380
QY 1381 AACCTCAAGAGATGATGATTTTGGTCTCAGTTGCGAGAGTCTCATTTAAAAATGCTAAA 1440
DB 1381 AACCTCAAGAGATGATGATTTTGGTCTCAGTTGCGAGAGTCTCATTTAAAAATGCTAAA 1440
QY 1441 ATAGATACAACTCAATTTKGSATWGAAGGGTWTTTTGRATWGYCCCATGTTGGGT 1500
DB 1441 ATAGATACAACTCAATTTKGSATWGAAGGGTWTTTTGRATWGYCCCATGTTGGGT 1500
QY 1501 CTCACATTAGAGTAATTTCTATTNAANCATGAAATTCGCTTTTATGAGTGATACCAAT 1560
DB 1501 CTCACATTAGAGTAATTTCTATTNAANCATGAAATTCGCTTTTATGAGTGATACCAAT 1560
QY 1561 ANGCGCTCTAANCCTTCATTTGNACTCAGTGAAGAGGGAAGCGGGAGAGTAATTT 1620
DB 1561 ANGCGCTCTAANCCTTCATTTGNACTCAGTGAAGAGGGAAGCGGGAGAGTAATTT 1620
QY 1621 NTTATGCTGAATGCGAGGATATTGGTCTGACTTACCNTAGGGGANTTTAAAACTGNC 1680
DB 1621 NTTATGCTGAATGCGAGGATATTGGTCTGACTTACCNTAGGGGANTTTAAAACTGNC 1680
QY 1681 CTTTTGAATCTGTTTGGATGCGCCTT 1707
DB 1681 CTTTTGAATCTGTTTGGATGCGCCTT 1707
RESULT 2
US-09-804-006-13
: Sequence 13, Application US/09804006
: Patent No. US20020119517A1
: GENERAL INFORMATION:
: APPLICANT: White, David
: APPLICANT: Zhou, Jiahong
: APPLICANT: Tartaglia, Louis A.
: TITLE OF INVENTION: LEPTIN INDUCED GENES
: FILE REFERENCE: 07334/126001
: CURRENT APPLICATION NUMBER: US/09/804,006
: PRIOR FILING DATE: 2001-03-12
: PRIOR APPLICATION NUMBER: US 09/292,228
: PRIOR FILING DATE: 1999-04-15
: PRIOR APPLICATION NUMBER: US 60/108,379
: PRIOR FILING DATE: 1998-10-29
: PRIOR APPLICATION NUMBER: US 09/150,857
: PRIOR FILING DATE: 1998-09-10
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 13
: LENGTH: 1707
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (246)...(1436)
: NAME/KEY: misc.feature
: LOCATION: (1)...(1707)
: OTHER INFORMATION: n = A,T,C or G
US-09-804-006-13

Query Match	99.3%;	Score 1694.8;	DB 10;	Length 1707;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1707;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	ACGCGTCCGCGAGCGCGCAGCGGCACAGCGGCAACAAGTGGCGGAGGCTAGCAGAGCCA	60
Db	1	ACGCGTCCGCGCAGCGCGCAGCGCAGCAGCGCAACAAGTGGCGGAGGCTAGCAGAGCCA	60
Qy	61	AGCGCGAGCAGTCCCTGCCGCGACACACCGCGCGGCCGCTCCCGGGCGCGCGCATGG	120
Db	61	AGCGCGAGCAGTCCCTGCCGCGACACCGCGGGGCCCGCTCCCGGGCGCGCGCATGG	120
Qy	121	AGCGTGAAGTGGCGGCTGCCCGGGTGAGCCGCGCGGAGCGCGCGAGCTGGATGTGG	180
Db	121	AGCGTGAAGTGGCGGCTGCCCGGGTGAGCCGCGCGGAGCGCGCGAGCTGGATGTGG	180
Qy	181	CGCGGATCTCCGCGCTTCCCGCCCGCCCGCGAGCTGGAGTGTCTCCCGGACAGATAT	240
Db	181	CGCGGATCTCCCGCGCTTCCCGCCCGCCCGCGAGCTGGAGTGTCTCCCGGACAGATAT	240
Qy	241	GAGAAATGAGTGTGACAGCTCGAAGAAATAAAGTTGTGGGTATCTCTGATGATGGCAATG	300
Db	241	GAGAAATGAGTGTGACAGCTCGAAGAAATAAAGTTGTGGGTATCTCTGATGATGGCAATG	300
Qy	301	TCTTCATTTATTTTATTTATTTAGAGTCTCCAAAAGCAGTAGCCAAAGAAAATGGAAG	360
Db	301	TCTTCATTTATTTTATTTATTTAGAGTCTCCAAAAGCAGTAGCCAAAGAAAATGGAAG	360
Qy	361	GGGAAGTAATAACCAAGAGAAATCTCGGAAGATATCTACCCCTCCCGAGGCATACT	420
Db	361	GGGAAGTAATAACCAAGAGAAATCTCGGAAGATATCTACCCCTCCCGAGGCATACT	420
Qy	421	GGAAACGAGCAGCAAGAGAAAGCTGAACGGGCGCTACAAACCCCATCTCTGAGCATGCTGACCA	480
Db	421	GGAAACGAGCAGCAAGAGAAAGCTGAACGGGCGCTACAAACCCCATCTCTGAGCATGCTGACCA	480
Qy	481	ACCAGAGGGGGAGCGGGGAGGCTCTCCAAATATAAGCCATCTGAACTACTGCGAACCTG	540
Db	481	ACCAGAGGGGGAGCGGGGAGGCTCTCCAAATATAAGCCATCTGAACTACTGCGAACCTG	540
Qy	541	ACCTGAGGCTCAGCTCGGTGGTTACGGTITTTAAACAATCTGCGGACAGATTTAAAGACT	600
Db	541	ACCTGAGGCTCAGCTCGGTGGTTACGGTITTTAAACAATCTGCGGACAGATTTAAAGACT	600
Qy	601	TTCTGCTGATTTGAGATGCCGAATTTATTCAGTCTTATAGATCAGCGGGATAGGTGTG	660
Db	601	TTCTGCTGATTTGAGATGCCGAATTTATTCAGTCTTATAGATCAGCGGGATAGGTGTG	660
Qy	661	CAAGAAGAACCTTTCTTTGCTGGGATTAAGTCCCTCACTCCACATTTTGCAGAGGC	720
Db	661	CAAGAAGAACCTTTCTTTGCTGGGATTAAGTCCCTCACTCCACATTTTGCAGAGGC	720
Qy	721	AAGCAATCCGGGAATCTTGGGGCAAGAAAGCAACGAGGGAACCAACCGTGGTGGAG	780
Db	721	AAGCAATCCGGGAATCTTGGGGCAAGAAAGCAACGAGGGAACCAACCGTGGTGGAG	780
Qy	781	TCTTCTGCTGGCCAGACACCCCGCAGAGCAACCAACCCCGACCTTTCAGATATGCTGA	840
Db	781	TCTTCTGCTGGCCAGACACCCCGCAGAGCAACCAACCCCGACCTTTCAGATATGCTGA	840
Qy	841	AATTTGAGAGTGAGAGCAACCAAGACATTTATGTTGGAACCTACAGAGACACTTCTTCA	900
Db	841	AATTTGAGAGTGAGAGCAACCAAGACATTTATGTTGGAACCTACAGAGACACTTCTTCA	900
Qy	901	ACTTGTCTCTGAAGGAGTCTGTTTCTCAGGTGGGTAACTCTTCTGCCAGACACTG	960
Db	901	ACTTGTCTCTGAAGGAGTCTGTTTCTCAGGTGGGTAACTCTTCTGCCAGACACTG	960
Qy	961	AGTTTGTGTTTCAAGGGCGATGACGATGTTTTTGTGAACCCCATCATCCTGAAATTA	1020
Db	961	AGTTTGTGTTTCAAGGGCGATGACGATGTTTTTGTGAACCCCATCATCCTGAAATTA	1020
Qy	1021	TGAATAGTTATTCGAAGACCAAAAGCAAGATCTCTTCATAGTGTGATGATGCACCAATG	1080

Db	1021	TGAATAGTTATCCAAGACCAAAAGACTCTCTTCATAGTGTGATCCACAATG	1080
Qy	1081	CTGAGCCTCATCGGATAGAAGCTGAAGTACTACATCCACAGAGTGTTTACTCTGGCC	1140
Db	1081	CTGAGCCTCATCGGATAGAAGCTGAAGTACTACATCCACAGAGTGTTTACTCTGGCC	1140
Qy	1141	TCTACCCACCTATGACGGGGGAGGGGGTTCCTCTACTCCGGCCACTCGCCCTGAGGC	1200
Db	1141	TCTACCCACCTATGACGGGGGAGGGGGTTCCTCTACTCCGGCCACTCGCCCTGAGGC	1200
Qy	1201	TGTACCATATCACTGACAGGTCATCTCTACCCCATTTGATGAGCTTACTTGGAAATGT	1260
Db	1201	TGTACCATATCACTGACAGGTCATCTCTACCCCATTTGATGAGCTTACTTGGAAATGT	1260
Qy	1261	GCCTTCAGAAACTCGGCCTCGTTCAGAGAAAACACAAAGGCTTCAGGACATTTGATATCG	1320
Db	1261	GCCTTCAGAAACTCGGCCTCGTTCAGAGAAAACACAAAGGCTTCAGGACATTTGATATCG	1320
Qy	1321	AGGAGAAAACAAAATAACATCTGCTCCTATGTAGATCTGTATTTAGTACATAGTAGAA	1380
Db	1321	AGGAGAAAACAAAATAACATCTGCTCCTATGTAGATCTGTATTTAGTACATAGTAGAA	1380
Qy	1381	AACCTCAAGAGATGATGATATTGGTCTCAGTTGCAGAGTGCTCATTTAAAATGCTAAA	1440
Db	1381	AACCTCAAGAGATGATGATATTGGTCTCAGTTGCAGAGTGCTCATTTAAAATGCTAAA	1440
Qy	1441	ATAGATACAAACTCAATTTKGSATWRAAGGGGTWTTTGRAWGWYCCCATGTGGGGT	1500
Db	1441	ATAGATACAAACTCAATTTKGSATWRAAGGGGTWTTTGRAWGWYCCCATGTGGGGT	1500
Qy	1501	CTCACATTAGATGAATTTCTATTINANCATGAAATTCGCTTTATGAGTGATACCCATT	1560
Db	1501	CTCACATTAGATGAATTTCTATTINANCATGAAATTCGCTTTATGAGTGATACCCATT	1560
Qy	1561	ANGGCCTCTAANCCTTCATTGTGNACTCACGTGAAGAGGGAAGCGGGAAGGTAATTT	1620
Db	1561	ANGGCCTCTAANCCTTCATTGTGNACTCACGTGAAGAGGGAAGCGGGAAGGTAATTT	1620
Qy	1621	NTTTATGGTGAATGGCAGGATATTGGTCTGACTTACCGNTAGGGGANTTTAAAACCTGNC	1680
Db	1621	NTTTATGGTGAATGGCAGGATATTGGTCTGACTTACCGNTAGGGGANTTTAAAACCTGNC	1680
Qy	1681	CTTTTGAATCTGTTGGATGGCCCTT 1707	
Db	1681	CTTTTGAATCTGTTGGATGGCCCTT 1707	
RESULT 3			
US-10-109-563-3			
; Sequence 3, Application US/10109563			
; Publication No. US20020170075A1			
; GENERAL INFORMATION:			
; APPLICANT: Leviten, Michael W.			
; APPLICANT: Phillips, Russell			
; TITLE OF INVENTION: TRANSGENIC MICE COMPRISING			
; TITLE OF INVENTION: BETA-1.3-N-ACETYLACTOSAMINE DISRUPTIONS			
; FILE REFERENCE: R-021			
; CURRENT APPLICATION NUMBER: US/10/109,563			
; CURRENT FILING DATE: 2002-07-09			
; PRIOR APPLICATION NUMBER: US 60/280,706			
; PRIOR FILING DATE: 2001-03-29			
; NUMBER OF SEQ ID NOS: 6			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 3			
; LENGTH: 1831			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-10-109-563-3			

Query Match 89.7%; Score 1531.4; DB 13; Length 1831;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1585; Conservative 8; Mismatches 25; Indels 7;

QY 13 ACCGGCAGCGCAGGAGCGGCAACAAGTCCGGAGGCTACAGAGCCAGCCGAGCAGT 72
DB 3 AGCGCAGCGCAGCGGAGCGGCAACAAGTCCGGAGGCTACAGAGCCAGCCGAGCAGT 52
QY 73 CCCTCCGCGCAGACCGCGCGCGCGCGTCCGGGGCGCGCGCATGGAGGCTGAGCTGC 132
DB 63 CCCTCCGCGCAGACCGCGCGCGCGCGTCCGGGGCGCGCGCATGGAGGCTGAGCTGC 122
QY 133 GCGGTGCGCGCGGTGAGCGCGCGGAGCGCGCGCGGAGCGTGGATGTGGCGCGGATCTGCC 192
DB 123 GCGGTGCGCGCGGTGAGCGCGCGGAGCGCGCGGAGCGTGGATGTGGCGCGGATCTGCC 182
QY 193 GCGCTTGCGCGCGCGCGCGCGGAGCTGAGCTGCCCGGACAAAGATATGACAAATGAGTG 252
DB 183 GCGCTTGCGCGCGCGCGCGCGGAGCTGAGCTGCCCGGACAAAGATATGACAAATGAGTG 242
QY 253 TTGGAGCTCGAAGATAAAGTTGTTGGGTATCTTGATGATGCGCAATGCTTCAATTAIT 312
DB 243 TTGGAGCTCGAAGATAAAGTTGTTGGGTATCTTGATGATGCGCAATGCTTCAATTAIT 302
QY 313 TTATTATGGAAGTCTCCAAAGCAGTAGCCAAAGAAAATGGAAAGGGAAGTAATAA 372
DB 303 TTATTATGGAAGTCTCCAAAGCAGTAGCCAAAGAAAATGGAAAGGGAAGTAATAA 362
QY 373 TACCCAAAGAGAGTTCTGGAAGATATCTACCCCTCCCGGAGGCATCTGGAACCGAGAGC 432
DB 363 TACCCAAAGAGAGTTCTGGAAGATATCTACCCCTCCCGGAGGCATCTGGAACCGAGAGC 422
QY 433 AAGAGAAGCTGAACCGGAGTACAAACCCATCTGAGCATGCTGACCAACGAGAGGGGG 492
DB 423 AAGAGAAGCTGAACCGGAGTACAAACCCATCTGAGCATGCTGACCAACGAGAGGGGG 482
QY 493 AGCGGGGAGGCTCTCCAAATATAGCGATCTCAACTCTGGAACTCGACCTGAGGGTCA 552
DB 483 AGCGGGGAGGCTCTCCAAATATAGCGATCTCAACTCTGGAACTCGACCTGAGGGTCA 542
QY 553 CGTCGGTGTGTACGGGTITTAACAACCTTCCCGGACAGATTTAAAGACTTTCTGCTGTAIT 612
DB 543 CGTCGGTGTGTACGGGTITTAACAACCTTCCCGGACAGATTTAAAGACTTTCTGCTGTAIT 602
QY 613 TCAGATGCCGCAATATTACACTGCTATAGATCAGCGGATAGTGTCAAGAAACCTT 672
DB 603 TGAGATGCCGCAATATTACACTGCTATAGATCAGCGGATAGTGTCAAGAAACCTT 662
QY 673 TCTTGTGCTGGCGATTAAAGTCCCTCACTCCACATTTTCCAGAAAGCAAGCAATCCGGG 732
DB 663 TCTTGTGCTGGCGATTAAAGTCCCTCACTCCACATTTTCCAGAAAGCAAGCAATCCGGG 722
QY 733 AATCTGGGGCCCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 792
DB 723 AATCTGGGGCCCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 782
QY 793 GCAGACACCCCGAGAGGACACCCCGACCTTCAGATATCTGAAATTTGAGAGTG 852
DB 783 GCAGACACCCCGAGAGGACACCCCGACCTTCAGATATCTGAAATTTGAGAGTG 842
QY 853 AGAAGCACCAGACATTTTATGGAAGTACAGAGACACATTTCTTCAACTTGTCTCTCA 912
DB 843 AGAAGCACCAGACATTTTATGGAAGTACAGAGACACATTTCTTCAACTTGTCTCTCA 902
QY 913 AGAAGTGTGTTTCTCAGGTGGTAAGTACTTCTGCGGACAGACACTGAGTTGTTTCA 972
DB 903 AGAAGTGTGTTTCTCAGGTGGTAAGTACTTCTGCGGACAGACACTGAGTTGTTTCA 962
QY 973 AGGCGCATGACGATGTTTTTGTGAACACCCATCAGATCTGAAATTTGATAGTTAT 1032
DB 963 AGGCGCATGACGATGTTTTTGTGAACACCCATCAGATCTGAAATTTGATAGTTAT 1022
QY 1033 CCAAGACCAAAGCAAGATCTCTTCAAGTGTGATGATGATGATGATGATGATGATGATGATG 1092
DB 1023 CCAAGACCAAAGCAAGATCTCTTCAAGTGTGATGATGATGATGATGATGATGATGATGATC 1082

QY 1093 GGCAIAGAGAGTGAAGTACTACATCCAGAGCTTCTTACTTCTGCGCTCTCCACCCT 1152
DB 1083 GGATAGAGAGCTGAAGTACTACATCCAGAGCTTCTTACTTCTGCGCTCTCCACCCT 1142
QY 1153 ATCAGGGGAGGGGGTCTCTACTCCGGCCACCTGGCCCTGAGGCTGACCATATCA 1212
DB 1143 ATCAGGGGAGGGGGTCTCTACTCCGGCCACCTGGCCCTGAGGCTGACCATATCA 1202
QY 1213 CTGACAGGTCCATCTCTACCCCATTTGATGACGTTTATCTGGAATGTCCTTCAGAAC 1272
DB 1203 CTGACAGGTCCATCTCTACCCCATTTGATGACGTTTATCTGGAATGTCCTTCAGAAC 1262
QY 1273 TCGGCTCGTTCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGGAGAAACAA 1332
DB 1263 TCGGCTCGTTCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGGAGAAACAA 1322
QY 1333 AAAATACATCTGCTCTATGTAGTCTGATTTAGTATAGTAAACCTCAAGAA 1392
DB 1323 AAAATACATCTGCTCTATGTAGTCTGATTTAGTATAGTAAACCTCAAGAA 1382
QY 1393 TGATTGATATTTGCTCTCAGTTCAGAGTCTCATTTAAATGCTAAATAGATACAAAC 1452
DB 1383 TGATTGATATTTGCTCTCAGTTCAGTTCAGAGTCTCATTTAAATGCTAAATAGATACAAAC 1442
QY 1453 TCAATTTGATGAGAGGGTWTGTTGATGAGGCTGATTTGATGAGTATGAGTATGAG 1512
DB 1443 TCAATTTGATGAGAGGGTWTGTTGATGAGTATGAGTATGAGTATGAGTATGAG 1502
QY 1513 TAATTTCTATTNNANC--ATGAATGCTTTATGATGATACCCATTT-ANGGCTCT 1569
DB 1503 TAATTTCTATTNNANC--ATGAATGCTTTATGATGATACCCATTT-ANGGCTCT 1562
QY 1570 AANCCIT--CATTTGNACTCAGTGAAGAAAGCGGAGAGGAGTAAATTTTAT 1626
DB 1563 AANCCITCAATTTGCTACTCAGTGAAG-AGGAAAGCGGAGAGTAAATTTTAT 1621
QY 1627 GGTGA 1631
DB 1622 ATGGA 1626

RESULT 4
US-09-972-912-1
: Sequence 1, Application US/09972912
: Patent No. US2002110867A1
: GENERAL INFORMATION:
: APPLICANT: SOPPET, DANIEL R.
: RUBEN, STEVEN M.
: TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
: STREET: 1100 NEW YORK AVENUE, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: US
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/972,912
: FILING DATE: 10-Oct-2001
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/049,022
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: STEFFE, ERIC K.
: REGISTRATION NUMBER: 36,688
: REFERENCE/DOCKET NUMBER: 1488.0620001

```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2745 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 233..1423
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 233..328
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 329..1423
: SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-972-912-1

Query Match      87.2%; Score 1488; DB 10; Length 2745;
Best Local Similarity 97.3%; P-red. No. 0;
Matches 1575; Conservative 8; Mismatches 24; Indels 11; Gaps 7;

QY 21 CGGACAGCGGCAACAAAGTGGCGGAGGCTAGCAGAGCCAGCCGAGCAGTCCCTGGCG 80
DB 10 CGGACAGCGGCAACAAAGTGGCGGAGGCTAGCAGAGCCAGCCGAGCAGTCCCTGGCG 67
QY 81 CGGACAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 140
DB 68 CGGACAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 127
QY 141 CC-GG8CTGAGCGCGCGCGGAGCGCGCGGACGTGGATGTGGCGGCGATCTCCGCGCCCTG 199
DB 128 CCGGGGCTCAGCGCGCGCGGAGC-GCGGGACGTGGATGTGGCGGCGATCTCCGCGCCCTG 186
QY 200 CCGCGCGCGCGCGGAGCTGGAGCTGCTCCGCGACAGATATGAGAAATGAGTGTGGAGG 259
DB 187 CCGCGCGCGCGCGGAGCTGGAGCTGCTCCGCGACAGATATGAGAAATGAGTGTGGAGG 246
QY 260 TCGAAGATTAAGTCTTGGGTATCTGATGATGCAATGCTTCATTTATTTATTTAT 319
DB 247 TCGAAGATTAAGTCTTGGGTATCTGATGATGCAATGCTTCATTTATTTATTTAT 306
QY 320 GGAAGTCTCCAAAGCAGTAGCCAGAGAAAAAATGGAAGAGGGAAGTAAATACCCCAA 379
DB 307 GGAAGTCTCCAAAGCAGTAGCCAGAGAAAAAATGGAAGAGGGAAGTAAATACCCCAA 366
QY 380 AGAAGAGTCTGGAAGATATCTACCCCTCCGAGGCTACTGGAACCGAGAGCAAGAGNA 439
DB 367 AGAAGAGTCTGGAAGATATCTACCCCTCCGAGGCTACTGGAACCGAGAGCAAGAGNA 426
QY 440 GCTGAACCGGCGAGTACAAACCCATCTGAGCATGCTGACCAACAGGAGGAGGAGCGGG 499
DB 427 GCTGAACCGGCGAGTACAAACCCATCTGAGCATGCTGACCAACAGGAGGAGGAGCGGG 486
QY 500 CAGGCTCTCCATATAGCATCTGAATCTACTGCGAACCTGACCTGAGGAGTCACTCGGT 559
DB 487 CAGGCTCTCCATATAGCATCTGAATCTACTGCGAACCTGACCTGAGGAGTCACTCGGT 546
QY 560 GGTACGGGTTTTAAACATTCGCGACAGATTTAAAGACTTCTGCTATTTAGAGATG 619
DB 547 GGTACGGGTTTTAAACATTCGCGACAGATTTAAAGACTTCTGCTATTTAGAGATG 606
QY 620 CCACAATTATTCATCTCTATAGATCAGCGGATAGTGTGCAAGAAACCTTTCTTGTT 679
DB 607 CCACAATTATTCATCTCTATAGATCAGCGGATAGTGTGCAAGAAACCTTTCTTGTT 666
QY 680 GCTGGCGATTAGTCCCTCACTCCACATTTTCCAGAGGCAAGCAATCCGGGAATCCGTG 739
DB 667 GCTGGCGATTAGTCCCTCACTCCACATTTTCCAGAGGCAAGCAATCCGGGAATCCGTG 726
```

RESULT 5

US-10-109-563-1

; Sequence 1, Application US/10109563

; Publication No. US20020170075A1

; GENERAL INFORMATION:

; APPLICANT: Leviten, Michael W.

; APPLICANT: Phillips, Russell

; TITLE OF INVENTION: TRANSGENIC MICE COMPRISING

; FILE REFERENCE: BETA-1,3-N-ACETYL LACTOSAMINE DISRUPTIONS

; CURRENT APPLICATION NUMBER: US/10/109,563

```

QY 740 GGGCCAAAGAAAGCAACACGAGGAAACCAACGGTGGTGGAGTCTTCTGCTGGGCCAGAC 799
DB 727 GGGCCAAAGAAAGCAACACGAGGAAACCAACGGTGGTGGAGTCTTCTGCTGGGCCAGAC 786
QY 800 ACCCCAGAGGAGCAACACCCAGCCGACCTTTCAGATATGCTGAAATTTGAGAGTGAAGCA 859
DB 787 ACCCCAGAGGAGCAACACCCAGCCGACCTTTCAGATATGCTGAAATTTGAGAGTGAAGCA 846
QY 860 CCAAGACATTTTATGTGGAACCTACAGAGACACTTTCTTCAACTGTGCTCTGAAGGAAGT 919
DB 847 CCAAGACATTTTATGTGGAACCTACAGAGACACTTTCTTCAACTGTGCTCTGAAGGAAGT 906
QY 920 GCTGTTCTCAGTGGGTAAAGTACTTCTGCCAGACACTGAGTGTGTTTCAAGGGCGA 979
DB 907 GCTGTTCTCAGTGGGTAAAGTACTTCTGCCAGACACTGAGTGTGTTTCAAGGGCGA 966
QY 980 TGAGGATGTTTTTGTGAACCCCACTACATCTCAATTTACTTGAATAGTTTATCCAAAGAC 1039
DB 967 TGAGGATGTTTTTGTGAACCCCACTACATCTCAATTTACTTGAATAGTTTATCCAAAGAC 1026
QY 1040 CAAAGCCAAAGATCTCTTCATAGTGTGATGATCCACAATGCTGAGACCTCATCGGGATAA 1099
DB 1027 CAAAGCCAAAGATCTCTTCATAGTGTGATGATCCACAATGCTGAGACCTCATCGGGATAA 1086
QY 1100 GAAGCTGAAGTACTACATCCAGAAAGTGTGTTTACTCTGGCCCTTACCCACCTATGCGAG 1159
DB 1087 GAAGCTGAAGTACTACATCCAGAAAGTGTGTTTACTCTGGCCCTTACCCACCTATGCGAG 1146
QY 1160 GGGAGGGGGTTCCTCTACTCCGCCACCTGAGGCTGAGGCTGTACCATATCACTGACCA 1219
DB 1147 GGGAGGGGGTTCCTCTACTCCGCCACCTGAGGCTGAGGCTGTACCATATCACTGACCA 1206
QY 1220 GGTCCATCTCTACCCCAATTGATGACGTTTATPACTGGAATGTGCTTCAGAACTTCGCGCT 1279
DB 1207 GGTCCATCTCTACCCCAATTGATGACGTTTATPACTGGAATGTGCTTCAGAACTTCGCGCT 1266
QY 1280 GTTCCAGAGAAACACAAGGCTTCAGGACATTTGATATCAGGAGAAACCAAAATAA 1339
DB 1267 GTTCCAGAGAAACACAAGGCTTCAGGACATTTGATATCAGGAGAAACCAAAATAA 1326
QY 1340 CATCTGCTCTATGATCTGATGTTAGTACATAGTAGAAGGCTCAAGAGATGATTGA 1399
DB 1327 CATCTGCTCTATGATCTGATGTTAGTACATAGTAGAAGGCTCAAGAGATGATTGA 1386
QY 1400 TATTGTGCTCAGTGTGAGAGTGTCTCATTTAAATGCTTAAATAGATACAAACTCAATTT 1459
DB 1387 TATTGTGCTCAGTGTGAGAGTGTCTCATTTAAATGCTTAAATAGATACAAACTCAATTT 1446
QY 1460 KGSATWGAAGGGTWTGTTTGRATWGGYCCCATGTTGGGCTCCACATAGAGTAATTTTC 1519
DB 1447 TGCAATAGAAAGTGTATTTTGAATAGTTCCTCATGTTGTTCTCAGATAGAGTAATTC 1506
QY 1520 TATTNRAAC--ATGAATTCGCTTTTATGATGATACCCATTT--ANGGCTCTAANCCTT 1576
DB 1507 TATATTAACCATGAAAAATGCTTTTATGATGATACCCATTTTTCAGGCGCTCTAAGCCT 1566
QY 1577 ---CATTTGNACTCAGTGAAGAGGGAAGCGGAGAGGAGTAATTTTATGTTGA 1631
DB 1567 TCAATTTGGTACTCAGCTGAAG-AGGGAAAGCGGAAGATGTTAATTTTTTTTATGA 1623
```



```
QY 484 AGACGGGGAGGGGGGAGGCTCTCCCAATATAAGCACTGAACTACTGCGAACCTGACC 543
Db 241 AGACAGGGAGCTAGCCACATCTCCAAACACAACTCACTGAGCTATTTGTGAACAGACT 300
QY 544 TGAGGGTCAGCTCGGTGTTACGGGTTTAAACAATGCGCGGACAGATTTAAAGACTTC 603
Db 301 CGACGGTCATCAGACGTGTCAGAGATTTTAAATATCTGCGGACAGATTTAAAGACTTC 360
QY 604 TGCTGTATTTAGATGCGGCAATTAATTCATCTGCTTATAGATACCGGATTAAGTGTCAA 663
Db 361 TCTTGTATTTAGATGCGGCAATTAATTCATCTGCTTATAGATACCGGATTAAGTGTCAA 420
QY 664 AGAAACCTTCTTCTGCTGCGATTAAGTCTCACTCCACATTTTGCCAGAGGCAAG 723
Db 421 AGAAGCCCTTCTTACTATTGGGCAATTAAGTCTCACTCCACATTTTGCCAGAGGCAAG 480
QY 724 CAATCCGGGAATCTTGGGGCCAGAAAGCAACCAAGGAGGAGGAGGAGGAGGAGGAG 783
Db 481 CAATCCGGGAATCTTGGGGCCAGAAAGCAACCAAGGAGGAGGAGGAGGAGGAGGAG 540
QY 844 TTGAGAGTGAAGACCAACCAAGCAATTTCTATGCTGCTTATAGATACCGGATTAAGTGTCAA 663
Db 601 TTGAGAGTGAAGACCAACCAAGCAATTTCTATGCTGCTTATAGATACCGGATTAAGTGTCAA 600
QY 904 TGCTGTATTTAGATGCGGCAATTAATTCATCTGCTTATAGATACCGGATTAAGTGTCAA 963
Db 661 TGCTGTATTTAGATGCGGCAATTAATTCATCTGCTTATAGATACCGGATTAAGTGTCAA 600
QY 964 TTGAGAGTGAAGACCAACCAAGCAATTTCTATGCTGCTTATAGATACCGGATTAAGTGTCAA 1023
Db 721 TTGAGAGTGAAGACCAACCAAGCAATTTCTATGCTGCTTATAGATACCGGATTAAGTGTCAA 780
```

RESULT 7

US-09-804-006-1

; Sequence 1, Application US/0984006

; Patent No. US2002011951/A1

; GENERAL INFORMATION:

; APPLICANT: White, David

; APPLICANT: Zhou, Jianghong

; APPLICANT: Tartaglia, Louis A.

```
QY 1024 ATAGTTTATCCAGACCAAGCCAAAGATCTCTTCATAGTGTATGATCCACAATGCTG 1063
DB 781 ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGTGTACGTGATCCACAATGCTG 840
QY 1084 GACCTCATCGGATAGAAAGCTGAAGTACTACATCCCAAGCTTGTACTCTGGCTCT 1143
DB 841 GGCCTACCGGATAGAAAGCTGAAGTACTACATCCCAAGCTTGTACTACCGGCTCT 900
QY 1144 ACCACCTTATCGAGGGGAGGGGTTCTCTACTCCGGCCACCTGGCCCTGAGGCTGT 1203
DB 901 ACCACCTTATCGAGGGGAGGGGTTCTCTACTCCGGCCACCTGGCCCTGAGGCTGT 960
QY 1204 ACCATATCACTGACAGGTCCTACTCTACGCCATCAIGACGCTTATAGTGAATGTGCC 1263
DB 961 ACAGTCGAGTACCGGGTCCATCTCTACCTCTATCTGATGTATATAGGGAATGTGCC 1020
QY 1264 TCCAGAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGG 1323
DB 1021 TCCAGAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATGAG 1080
QY 1324 AGAAACAAATAACATCTCTCTCTATGATCTGTATGATCTGTATGATAGATAAGAAAC 1383
DB 1081 AGAAACAAATAAGAAATAATTTGTTCTATATAGACTAATGTTAGTACATAGCAAAAC 1140
QY 1384 CTCAGAGATGATTGATATTTGTTCTCAGTTCGAGAGTCTCATTTAAATGCTAA 1439
DB 1141 CTCAGAGATGATTGATATTTGTTCTCAGTTCGAGAGTCTCATTTAAATGCTGA 1196

RESULT 8
US-09-804-357-3
; Sequence 3, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1996-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1191
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-804-357-3

Query Match 53.0%; Score 904.6; DB 9; Length 1191;
Best Local Similarity 85.0%; Pred. No. 7.1e+261;
Matches 1012; Conservative 0; Mismatches 179; Indels 0; Gaps 3;

QY 246 ATGAGTGTGGAGCTCGAAGATAAAGTTGTGGGTATCTCTGATGATGCAAAATGCTTTC 305
DB 1 ATGAGTGTGGGCGTCCAGAGTCAAGTTGCTGGCATCTCTGATGATGCAAAATGCTTTC 50
QY 306 ATTTATTTTATTTAGTCTCCAAAGCAGTAGCCAGCAAAATAATGAAAGGGGAA 365
DB 61 ATTTATTTGATTTGGAAGTCTCCAAAGCAGTAGCCAGCAAAATAATGAAAGGGGAA 120
QY 366 GTAAATATACCCAGAGAGTCTCTGGAAGATATCTACCCCTCCCGAGGCATCTGGAAC 425
DB 121 GTAAATATCCGAAAGAGAGTCTCTGGAAGCAGTCTCCCGAGGCATCTGGAAC 180
QY 426 CGAGAGCAAGAGAGTGAACCGGCGAGTACAAACCCCATCTCTGAGCATCTGACCAACCAG 485
```

```
DB 181 AGGCAACAGGAGAGCTGAACAGCTGTACAAATCCCATCTTGAACAGGTGGCCAAATCAG 240
QY 486 ACGGGGAGGGGGCGAGGCTCTCCAAATATAAGCATCTGAACCTACTCGCAACCTGACCTG 545
DB 241 ACGGGGAGCTAGCCACATCTCCAAACACAAAGTCCCTGAGCTATTGTGAACAGAGCTG 300
QY 546 AGGTCAGCTCGGTGGTTACGGGTTTTAACAACTTGGCGGACAGATTTAAAGACTTTTCG 605
DB 301 ACGTCTAGCAGAGCTGTGACAGATTTTAAATATCTGGCGGACAGATTTAAAGACTTTTC 360
QY 606 CTGTATTGAGATCCGCAATTTATTCATCTGCTTATAGATCAGCCGGAATAGTGTGCAAG 665
DB 361 TTGTATTGAGATCCGGAATTTACTCGCTTATAGATCAACCGAAGAATGTGCAAG 420
QY 666 AAACCTTTCTTTGTTGCTGGGATTAAGTCCCTACTCCATTTTCCAGATATGCTGAAATTT 725
DB 421 AAGCCCTTCTTACTATTGGCGATTAAGTCCCTCATTTCCACATTTTCCAGAAAGCAAGCA 480
QY 726 ATCCGGGATCTCTGGGCGCAAGAAAGCAACGAGGGAACCAACCGTGGTGGAGTCTTC 785
DB 481 ATCCGGGAGTCTTGGGCGGAGAAACCAACGTAAGGAAACGAGAGTAGTGAGGCTTTC 540
QY 786 CTGCTGGGCCAGACACCCCGAGAGCAACCAACCCGACCTTTCCAGATATGCTGAAATTT 845
DB 541 CTGTTGGGCAAGACCCCGAGAGCAACCAACCCGACCTTTCCGACATGCTTAAGTTT 600
QY 846 GAGAGTGAAGAGCAACCAAGACATTTCTATGTGGAACTACAGAGACACTTTCTTCAACTG 905
DB 601 GAGAGTGAAGAGCAACCAAGACATCTCATGTGGAATTAAGAGACACTTTCTTCAACTG 660
QY 906 TCTCTGAAGGAAGTCTGTTTCTCAGGTGGGTAACTTCTCTCCCGAGACACTGAGTTT 965
DB 661 TCCCTGAAGGAAGTCTGTTTCTTAGGTGGGTGAGCAGTCTCTCTCCAGCAGGAGTTT 720
QY 966 GTTTTCAAGGCGATGACGATGTTTTTTGTGAACCCCATCATCTCTGAAATTTACTTGAAT 1025
DB 721 GTCTTCAAGGCGATGATGAGTGTGTTGTGAACCCCATCATCTCTTAAATTTACTTGAAT 780
QY 1026 AGTTTATCAAGACCAAGCCAAAGATCTCTTCAATAGTGTATGATCCCAATGCTGGA 1085
DB 781 AGCTTATCAAGAGCAAAAGCCAAAGACTTGTTCATAGGTGACGTGATCCCAATGCTGGG 840
QY 1086 CTTATCGGGATTAAGAGCTGAAGTACTACATCCCAAGAGTTGTTTACTCTGGCTCTAC 1145
DB 841 CTTACCGGATTAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACGGGCTCTAC 900
QY 1146 CCACCTTATGAGGGGAGGGGTTCTCTTACTCCGGCCACCTGGCCCTGAGGCTGAC 1205
DB 901 CCACCTTATGAGGGGAGGGGTTCTCTTACTCCGGCCACCTGGCCCTGAGGCTGAC 960
QY 1206 CATATCACTGACAGGTCCTCTTACCCCATTCATGAGTGTATATCTGAAATGTCCTT 1265
DB 961 AGTGGAGTACCGGGTCCATCTCTACCTTATTGATGATGTTTATACGGGAATGTCCTT 1020
QY 1266 CAGAACTCGGCTCGTTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGAG 1325
DB 1021 CAGAACTCGGCTCGTTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATTTGAAG 1080
QY 1326 AAAACAAATAATACATCTGCTCTCTATGATGATCTGATGTTAGTACATAGTAAACCT 1385
DB 1081 AAAATAAGAAAAATATTTGTTCTATATAGACCTAATGTTAGTACATAGCAAAAACT 1140
QY 1386 CAAGAGATGATTGATTTTGTCTCAGTTGAGAGTGTCTCATTTTAAATGTC 1436
DB 1141 CAAGAGATGATTGATTTGTTCTGCTCAGTTGCAAGTCTTAATTTTAAATGTC 1191
```

```
RESULT 9
US-09-804-006-3
; Sequence 3, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
```

Qy	966	GT	TTTCAAGGGCGATGACGATGTTTTTTGTGAACACCCATCACATCTCTGAATTA	1020
Db	721	GT	CTTCAAGGGCGATGATGACGTTGTTGTGAACACCCATCACATCTCTGAAT	760
Qy	1026	AG	TTTATCCAAGACCAAGCCAAAGATCTCTTTCATAGGTGATGATCCACAATGCTGGA	1085
Db	781	AG	CTTATCCAAGACCAAGCCAAAGACTTCTTTCATAGGTGACGTGATCCACAATGCTGGG	840
Qy	1086	CT	CATCGGGATGAAGACTGAAGTACTACATCCCAAGAGTTGTTTACTCTGGCCTCTAC	1145
Db	841	CT	CACCGGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACGGCGCTCTAC	900
Qy	1146	CC	ACCTATCGAGGGGAGGGGGTTCCTCTACTCCGGCCACTGCGCCTGAGGCTGTAC	1205
Db	901	CC	ACCTATCGAGGGGAGGGGGTTCCTCTACTCCGGCCACTGCGCCTGAGGCTGTAC	960
Qy	1206	CA	TATCACTGACGAGGTCCATCTCTACCCCAATGTAGTACGCTTTATAGTGAATGTGCTT	1265
Db	961	AG	TGCGACTAGCGGGTCCATCTCTACCCCTATTGATGATCTTTATACGGGAATGTGCTT	1020
Qy	1266	CA	GAACCTCGGCTCGTTCAGAGAACACAAAGGCTTCAGGACATTTGATATCGAGGAG	1325
Db	1021	CA	GAACCTCGGCTCGTTCAGAGAACACAAAGGCTTCAGGACATTTGATATCGAGGAG	1080
Qy	1326	AAAA	CAAAATAACATCTCTCTATGTAGATCTGATGTTTAGTACATAGTAAACCT	1385
Db	1081	AAAA	TAAGAAAAATAATTTCTCTATAGACCTAATGTTTAGTACATAGTAAACCT	1140
Qy	1386	CA	GAGATGATGATATTTGGTCTCAGTTCGAGTGCTCTATTAAATGC	1436
Db	1141	CA	GAGATGATGATATCTCGTCTCAGTTCGAGTGCTCTAATTTAAATGC	1191

RESULT 10
 US-09-972-912-20
 : Sequence 20, Application US/09972912
 : Patent No. US20020110867A1
 : GENERAL INFORMATION:
 : APPLICANT: SOPPET, DANIEL R.
 : RUBEN, STEVEN M.
 : TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
 : NUMBER OF SEQUENCES: 42
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
 : STREET: 1100 NEW YORK AVENUE, SUITE 600
 : CITY: WASHINGTON
 : STATE: DC
 : COUNTRY: US
 : ZIP: 20005-3934
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/972,912
 : FILING DATE: 10-Oct-2001
 : CLASSIFICATION: <Unknown>
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 09/049,022
 : FILING DATE: <Unknown>
 : ATTORNEY/AGENT INFORMATION:
 : NAME: STEFFE, ERIC K.
 : REGISTRATION NUMBER: 36,688
 : REFERENCE/DOCKET NUMBER: 1488-0620001
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (202) 371-2600
 : TELEFAX: (202) 371-2540
 : INFORMATION FOR SEQ ID NO: 20:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 857 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: both


```

;
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0620001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-09-972-912-26

Query Match          19.2%   Score 328.2;   DB 10;   Length 480;
Best Local Similarity 92.0%   Pred. No. 8.9e-88;
Matches 402;   Conservative 0;   Mismatches 25;   Indels 10;   Gaps 5;

QY 544 TGAGGGTCAGTCGGTGGTTACGGGTTTAAACAACCTTGC CGGACAGATTAAAGACTTTC 603
Db 1 TGAGGGTCAGTCGGTGGTTACGGGTTTAAACAACCTTGC CGGACAGATTAAAGACTTTC 60
QY 604 TGTGTATTGTAGATGCCGAATTTATTCACCTGCTTATAGATCAGCCGGATAGGTGCGAA 663
Db 62 TGTGTATTGTAGATGCCGAATTTATTCACCTGCTTATAGATCAGCCGGATAGGTGCGAA 120
QY 664 AGAAACCTTCTTGTGCTGGCGATTAAAGTCCCTACTCCACATTTTGC CAGAGGCAA - 722
Db 121 AGAAACCTTCTTGTGCTGGCGATTAAAGTCCCTACTCCACATTTTGC CAGAGGCAAAG 180
QY 723 GCAATCGGGAATCCTGGGCCAAGAAAGCAACGCCAGGGAACCAACCGTGGTGGCGAGTC 782
Db 181 GCAATCGGGAATCCTGGGCCAAGAAAGCAACGCCAGGGAACCAACCGTGGTGGCGAGTC 240
QY 783 TTCCTGCTGGGCCAGACACCCCGCAGAGGAGCAACACCCCGACCTTTTCAGATATGCTGAAA 842
Db 241 TTCCTGCTGGGCCAGACACCCCGCAGAGGAGCAACACCCCGACCTTTTCAGATATGCTGAAA 300
QY 843 -TTTGAGAGTGAGAGCAACCAAGACATTTATATGTGGAA---CTACAGAGACACTTTC 898
Db 301 -TTTGAGAGTGAGAGCAACCAAGACATTTATATGTGGAA---CTACAGAGACACTTTC 360
QY 899 CAACCTGTCTCT-CAAGGAAGTGTCTTCT- - - -CAGGTGGGTAAAGTACTTCTCTGCCCA 953
Db 361 CAANTTGTCTNTGGAAGAGTGTCTGTTTTTTCAGGTGGGTAAAGTATTTCTCTGCCCA 420
QY 954 GACACTGAGTTTGTGTTT 970
Db 421 GACATGAGTTTGTGTTT 437

RESULT 12
US-09-972-912-31
; Sequence 31, Application US/09972912
; Patent No. US20020110867A1
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R.
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

;
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0620001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-972-912-20

Query Match          25.3%   Score 431.4;   DB 10;   Length 857;
Best Local Similarity 92.9%   Pred. No. 1.2e-118;
Matches 466;   Conservative 8;   Mismatches 25;   Indels 3;   Gaps 3;

QY 103C TATCCAAAGACCAAGCAAGATCTCTTCATAGGTGATGATCCACAAATGCTGGACCTC 1089
Db 20 TATCCAAAGACCAAGCAAGATCTCTTCATAGGTGATGATCCACAAATGCTGGACCTC 79
QY 1090 ATCCGGATGAAGAGCTGAAGTACTACATCCAGAGAGTTGTTACTCTGCGCTCTACCCAC 1149
Db 80 ATCCGGATGAAGAGCTGAAGTACTACATCCAGAGAGTTGTTACTCTGCGCTCTACCCAC 139
QY 1150 CCTATCGAGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGCTGTACCATTA 1209
Db 140 CCTATCGAGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGCTGTACCATTA 199
QY 1210 TCACCTGACAGCTCCATCTCTACCCCATTTGATGAGCTTATACGTGAATGCGCTTCAGA 1269
Db 200 TCACCTGACAGCTCCATCTCTACCCCATTTGATGAGCTTATACGTGAATGCGCTTCAGA 259
QY 1270 AACTCGGCTCGTCTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGGAGAAA 1329
Db 260 AACTCGGCTCGTCTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGGAGAAA 319
QY 1330 ACAAAATAACATCTGCTCTATGTAGATCTCATGATCTAGTACATAGTA-GAAACCTCAA 1388
Db 320 ACAAAATAACATCTGCTCTATGTAGATCTCATGATCTAGTACATAGTA-GAAACCTCAA 379
QY 1389 GAGATGATTGATTGTGCTC-AGTTCAGAGTGTCTC-ATTTAAATGCTAAATAGAT 1446
Db 380 GAGATGATTGATTGTGCTC-AGTTCAGAGTGTCTC-ATTTAAATGCTAAATAGAT 439
QY 1447 ACAAACTCAATTKGATNGRAGGGGTTTGTGATNGVCCCATGTGTGGGCTTCACA 1506
Db 440 ACAAACTCAATTKGATNGRAGGGGTTTGTGATNGVCCCATGTGTGGGCTTCACA 499
QY 1507 TTAGAGTAATTTCTATTNANCA 1530
Db 500 TTAGAGNGGTTCAAGTGGGTACA 523

RESULT 11
US-09-972-912-26
; Sequence 26, Application US/09972912
; Patent No. US20020110867A1
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R.
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/972,912
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/049,022
```


COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/972.912
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/049,022
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0620001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:

LENGTH: 303 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: both

MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 31:

US-09-972-912-31

Query Match 16.7%; Score 284.8; DB 10; Length 303;
Best Local Similarity 99.3%; Pred. No. 7.2e-75;
Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	585	GACAGATTAAAGACTTTCGCTGCTATTGAGATGCCGCAATATTCACCTGCTTAGAT	644
Db	1	GACAGATTAAAGACTTTCGCTGCTATTGAGATGCCGCAATATTCACCTGCTTAGAT	60
Qy	645	CAGCCGGATAGTGTGCAAGAAACCTTCTGTTGCTGGCGATTAAAGTCCCTCACTCCA	704
Db	61	CAGCCGGATAGTGTGCAAGAAACCTTCTGTTGCTGGCGATTAAAGTCCCTCACTCCA	120
Qy	705	CAITTTCCGAGAGGCAAGCAATCCGGGAATCCTGCGGGCCAAAGAACGACGAGGAAC	764
Db	121	CAITTTCCGAGAGGCAAGCAATCCGGGAATCCTGCGGGCCAAAGAACGACGAGGAAC	180
Qy	765	CAACGGTGTGGCAGTCTCTCTGCGGCAGACACCCCGAGAGCAACCCCGAC	824
Db	181	CAACGGTGTGGCAGTCTCTCTGCGGCAGACACCCCGAGAGCAACCCCGAC	240
Qy	825	CTTTCAGATATGCTGAAATTGAGAGTGAGAGCAACCAACATCTTCT	872
Db	241	CTTTCAGATATGCTGAAATTGAGAGTTAGAGCAACCAACATCTTCT	288

RESULT 13

US-09-972-912-11

Sequence 11, Application US/09972912
Patent No. US20020110867A1

GENERAL INFORMATION:

APPLICANT: SOPPET, DANIEL R.

RUBEN, STEVEN M.

TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

STREET: 1100 NEW YORK AVENUE, SUITE 600

CITY: WASHINGTON

STATE: DC

COUNTRY: US

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/972,912
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/049,022
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.

REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0620001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 282 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: both

MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 11:

US-09-972-912-11

Query Match 16.3%; Score 279; DB 10; Length 282;
Best Local Similarity 98.9%; Pred. No. 3.8e-73;
Matches 279; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy	707	TTTGGCAGAGGCAAGCAATCCGGGAATCCTGCGGGCCAAAGAACGACGAGGAACCA	766
Db	1	TTTGGCAGAGGCAAGCAATCCGGGAATCCTGCGGGCCAAAGAACGACGAGGAACCA	60
Qy	767	ACGGTGTGGCAGTCTCTCTGCGGCAGACACCCCGAGAGCAACCCCGACCT	826
Db	61	ACGGTGTGGCAGTCTCTCTGCGGCAGACACCCCGAGAGCAACCCCGACCT	120
Qy	827	TCAGATATGCTGAAATTTGAGAGTGAGAGCAACCAAGACATTTCTTATGTGAACCTACAG	886
Db	121	TCAGATATGCTGAAATTTGAGAGTGAGAGCAACCAAGACATTTCTTATGTGAACCTACAG	180
Qy	887	AGACACTTTCTCAACTGCTCTCTGAGGAAGTGTCTTCTCAGTGGGTGAAGTCTTC	946
Db	181	AGACACTTTCTCAACTGCTCTCTGAGGAAGTGTCTTCTCAGTGGGTGAAGTCTTC	240
Qy	947	CTGCCAGACACTGATTTTCTTCAAGGGCGATGACGATGT	988
Db	241	CTGCCAGACACTGATTTTCTTCAAGGGCGATGACGATGT	282

RESULT 14

US-09-972-912-34

Sequence 34, Application US/09972912
Patent No. US20020110867A1

GENERAL INFORMATION:

APPLICANT: SOPPET, DANIEL R.

RUBEN, STEVEN M.

TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

STREET: 1100 NEW YORK AVENUE, SUITE 600

CITY: WASHINGTON

STATE: DC

COUNTRY: US

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/972,912

FILING DATE: 10-Oct-2001

```
? CLASSIFICATION: <Unknown>
? PRIOR APPLICATION DATA:
? FILING DATE: <Unknown>
? ATTORNEY/AGENT INFORMATION:
? NAME: STEFFE, ERIC K.
? REGISTRATION NUMBER: 36,688
? REFERENCE/DOCKET NUMBER: 1488.0620001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 371-2600
? TELEFAX: (202) 371-2540
? INFORMATION FOR SEQ ID NO: 34:
? LENGTH: 282 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: both
? MOLECULE TYPE: cdna
? SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-972-912-34

Query Match      16.3%; Score 279; DB 10; Length 282;
Best Local Similarity 98.9%; Pred. No. 3.8e-73;
Matches 279; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 707 TTTGCGAGAGCAAGCAATCCGGGAATCCTGGGCCAAGAAAGCAAGCAGGGAACCA 766
Db 1 TTTGCGAGAGCAAGCAATCCGGGAATCCTGGGCCAAGAAAGCAAGCAGGGAACCA 60

QY 767 AACGGTGGTGGAGTCTTCCTGCTGGCCAGACACCCAGAGAGCAACCCCGACCT 826
Db 61 AACGGTGGTGGAGTCTTCCTGCTGGCCAGACACCCCGAGAGCAACCCCGACCT 120

QY 827 TTCAGATATGCTGAATTTGAGAGTGAAGAGCAAGCAAGCAATCTTAITGTGGAACACAG 886
Db 121 TTCAGATATGCTGAATTTGAGAGTGAAGAGCAAGCAAGCAATCTTAITGTGGAACACAG 180

QY 887 AGACACTTTCTCAACTTG-CTCTGAAGGAAGTCTGTTCTCAGTGGGTAAAGTACTC 946
Db 181 AGACACTTTCTCAACTTGCTGTAAGGAAGTCTGTTCTCAGTGGGTAAAGTACTC 240

QY 947 CTGCCAGACACTCAGTCTTTTCAAGGGCGATGACGATGT 988
Db 241 CTGCCAGACACTGAGTTTGTTTCAAGGGCGATGACGATGT 282

RESULT 15
US-09-972-912-36
? Sequence 36, Application US/0972912
? Patent No. US20020110867A1
? GENERAL INFORMATION:
? APPLICANT: SOPPET, DANIEL R.
? RUBEN, STEVEN M.
? TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
? NUMBER OF SEQUENCES: 42
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
? STREET: 1100 NEW YORK AVENUE, SUITE 600
? CITY: WASHINGTON
? STATE: DC
? COUNTRY: US
? ZIP: 20005-3934
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/972,912
? FILING DATE: 10-Oct-2001
? CLASSIFICATION: <unknown>
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 09/049,022
```

```
? FILING DATE: <Unknown>
? ATTORNEY/AGENT INFORMATION:
? NAME: STEFFE, ERIC K.
? REGISTRATION NUMBER: 36,688
? REFERENCE/DOCKET NUMBER: 1488.0620001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 371-2600
? TELEFAX: (202) 371-2540
? INFORMATION FOR SEQ ID NO: 36:
? LENGTH: 428 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: both
? MOLECULE TYPE: cdna
? SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-972-912-36

Query Match      15.7%; Score 267.4; DB 10; Length 428;
Best Local Similarity 99.6%; Pred. No. 1.6e-69;
Matches 268; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 942 ACTTCTGCCCCAGACACTGAGTTTGTTCRAAGGGCGATGACGATGTTTTCGTGAACACC 1001
Db 1 ACTTCTGCCCCAGACACTGAGTTTGTTCRAAGGGCGATGACGATGTTTTCGTGAACACC 60

QY 1002 CATCATCTCCTGAATTACTTGAATAGTTTATCCAGACCAAAAGCCAAAGATCTCTTCATA 1061
Db 61 CATCATCTCCTGAATTACTTGAATAGTTTATCCAGACCAAAAGCCAAAGATCTCTTCATA 120

QY 1062 GGTGATGATCCCAATGCTGGACCTCATCGGGATAGAAGCTGAAGTACTACATCCCA 1121
Db 121 GGTGATGATCCCAATGCTGGACCTCATCGGGATAGAAGCTGAAGTACTACATCCCA 180

QY 1122 GAAGTGTGTTTACTCTGGCCTCTACCCACCTATGCAGGGGGAGGGGTTCTCTACTCC 1181
Db 181 GAAGTGTGTTTACTCTGGCCTCTACCCACCTATGCAGGGGGAGGGGTTCTCTACTCC 240

QY 1182 GCCACCTGGCCCTGAGGCTGTACCATAT 1210
Db 241 GCCACCTGGCCTGAGGCTGTACCATAT 269

Search completed: October 20, 2003, 22:44:25
Job time : 473.998 secs
```